

Sharing AIS-Related Anomalies (SARA)

Dan Radulescu
Marie-Odette St-Hilaire
Yannick Allard
Tim R. Hammond
OODA Technologies Inc.

Prepared By:
OODA Technologies Inc.
4891 Grosvenor
Montreal, QC H3W 2M2

PWGSC Contract Number: W7707-145677/001/HAL
Technical Authority: Tim R. Hammond

Disclaimer: The scientific or technical validity of this Contract Report is entirely the responsibility of the Contractor and the contents do not necessarily have the approval or endorsement of the Department of National Defence of Canada.

Contract Report
DRDC-RDDC-2016-C182
March 2016

- © Her Majesty the Queen in Right of Canada, as represented by the Minister of National Defence, 2016
- © Sa Majesté la Reine (en droit du Canada), telle que représentée par le ministre de la Défense nationale, 2016

SHARING AIS-RELATED ANOMALIES (SARA)



Dan Radulescu
Marie-Odette St-Hilaire
Yannick Allard
Tim R. Hammond

Prepared By: OODA Technologies Inc.
4891 Grosvenor
Montréal (Qc), H3W 2M2
514.476.4773

Prepared For: Defence Research & Development Canada, Atlantic Research Centre
9 Grove Street, PO Box 1012
Dartmouth, NS
B2Y 3Z7
902-426-3100

Scientific Authority: Tim R. Hammond
Contract Number: W7707-145677/001/HAL
Call Up Number: 12
Project: Design, develop, manage, test and/or implement specific software or data source modules
Report Delivery Date: March 15, 2016

The scientific or technical validity of this Contract Report is entirely the responsibility of the contractor and the contents do not necessarily have the approval or endorsement of Defence R&D Canada.

This page is intentionally left blank.

Executive Summary

A diverse array of government and private organizations within Canada pursue the expanded use of the Automatic Identification System (AIS) for security purposes. These organizations include port authorities, police, coast guard and border security. In addition, Canada's allies are expected to adopt the standard, with just as broad a community of stakeholders. As a result, an extensive community is developing with a shared interest in the quality of AIS data.

AIS data quality is reflected in large part by the degree in which its application differs from its intended design. Though an AIS transponder is largely automated, opportunities exist for variability in practices, misconfiguration and intentional misuse. These unintended behaviours generate an abundance of anomalies that the security community has an interest in monitoring and sharing, especially in cases indicative of malicious intent.

The purpose of this project is to design a software system, called Sharing AIS-Related Anomalies (SARA), by which diverse users, who employ a variety of applications for viewing and storing AIS data, can share information about the AIS-related anomalies they uncover.

This document proposes a taxonomy and a representation of the metadata and information elements of the AIS-related anomalies for efficient discovery and sharing. It also includes a design description of SARA with use cases and requirements. Due to the potential involvement of different parties in the Canadian maritime security community, a survey of the AIS display and storage products in use within the community was conducted with the objective of identifying likely collaborators in the private sector. Finally, an application was produced to demonstrate the value of SARA and is described in this document.

This page is intentionally left blank.

Contents

Executive Summary	i
Contents	iii
List of Figures	ix
List of Tables	xiii
1 Introduction	1
2 Investigation of Collaboration Opportunities	3
2.1 Methodology	3
2.2 Canadian Marine Security Main Actors	4
2.3 AIS Display and Storage Main Products Used by the Canadian Community	4
2.3.1 ECPINS/SHINNADS	5
2.3.1.1 Clients	5
2.3.1.2 Integration	5
2.3.2 Interdepartmental Maritime Integrated Command, Control and Communi- cations (IMIC3)	6
2.3.2.1 Clients	6
2.3.2.2 Integration	6
2.3.3 GCCS-M	6
2.3.3.1 Clients	7
2.3.3.2 Integration	7

2.3.4	Halifax Port Authority Command and Control System	7
2.3.4.1	Clients	8
2.3.5	Provincial Aerospace	8
2.3.5.1	Clients	8
2.3.5.2	Integration	8
2.3.6	TimeCaster	9
2.3.6.1	Clients	9
2.3.6.2	Integration	9
2.3.6.3	Anomalies	9
2.3.7	exactEarth	9
2.3.7.1	Clients	10
2.3.7.2	Integration	10
2.3.7.3	Anomalies	10
2.3.8	Vessel Selection System	11
2.3.8.1	Clients	11
2.3.8.2	Integration	11
2.3.8.3	Anomalies	12
2.3.9	MSSIS-TV32/SeaVision	12
2.3.9.1	Clients	12
2.3.9.2	Integration	13
2.3.9.3	Anomalies	13
2.4	Potential Collaborators	13
3	Taxonomy of AIS-related Anomaly Types	15
3.1	AIS Anomalies Levels 1 and 2	15
3.2	Taxonomy	16
3.2.1	Error in AIS Messages (Level 1)	17
3.2.2	Transition from Anomalies Level 1 to 2	20

3.2.2.1	Time Dimension	24
3.2.3	Low Priority Anomalies	25
3.3	Maritime Kinematic Anomalies	25
4	Representation and Sharing of Metadata about Anomalies	27
4.1	National Information Exchange Model	27
4.1.1	Maritime Enterprise Information Exchange Model	29
4.2	NIEM Maritime Anomaly	32
4.3	Record and Security Metadata	33
4.4	NIEM-based Anomaly Exchange Report	34
4.5	Search Parameters for an Anomaly Report	36
4.6	Data Element Description	39
4.6.1	Vessel Identification	39
4.6.2	Anomaly Identification	40
4.6.3	Record Metadata	40
4.7	Suspicious Activity Reporting and its Lessons Learned	43
5	Requirements	45
5.1	External Interface Requirements	45
5.2	Software Interfaces	46
5.3	System Features	46
5.3.1	User Reporting of Anomalies	46
5.3.2	Retrieval of Anomalies	47
5.3.3	Retrieval and Edition of Ship Profiles	48
5.3.4	User profiles	48
5.4	Other System Features	49
5.4.1	Autonomous Generation of Anomaly Reports	49
5.4.2	Performance Tracker	50
5.4.3	Pre-processing watchdog	50

5.5	Other Nonfunctional Requirements	51
5.5.1	Design and Implementation Constraints	51
6	Design	53
6.1	Architectural Design	53
6.2	SARA Application Server	54
6.2.1	Anomaly Submission	56
6.2.2	Anomaly Retrieval	59
6.2.3	Ship Profile Editing	60
6.2.4	Ship Profile Retrieval	63
6.2.5	User Creation and Management	64
6.2.6	Authentication	66
6.3	Database and Data Structure	69
6.3.1	Vessel Profile Data Structure	70
6.3.2	Vessel Data Structure	70
6.3.3	Anomaly Data Structure	74
6.3.4	Contact Information Data Structure	74
6.3.5	Block Metadata Data Structure	78
6.3.6	Proposed Database Table Structure	78
6.3.7	SQL Versus NoSQL	81
6.4	Data Processing	83
6.4.1	Vessel Identification	85
6.4.2	Reliability Rating	85
6.4.3	Autonomous Generation of Anomaly Reports	86
6.4.4	Automatic Detection of Level 2 Anomalies from Level 1 Anomalies	86
6.5	Design Rationale	87
6.5.1	Service Architecture	87
6.5.2	Query Management	88

6.5.3	Authority Based User Creation	88
6.6	Integration into a NIEM Sharing Environment	88
7	Risks	91
7.1	Handling Ambiguities in Ship Identity	91
7.2	Creating Vessel Profiles from Anomaly Reports	92
7.3	Large Volume of Type 1 Anomalies	92
7.4	Integration of the Anomaly Retrieval Component	92
8	Demonstration Application	95
8.1	Scenario	95
8.2	Detailed Story	96
9	Conclusion	101
	Bibliography	103
A	Use cases	A-1
A.1	Uploading a Ship Anomaly	A-1
A.2	Retrieving Ship Anomalies from SARA	A-3
A.3	Retrieving Ship Profiles from SARA	A-5
A.4	Editing a Ship Profile for SARA	A-6
A.5	Creating a New User Profile in SARA	A-7
A.6	Editing an User Profile	A-8
B	Demonstration Application Installation	B-11
B.1	Installation and Launch	B-11
B.2	Troubleshooting	B-13

This page is intentionally left blank.

List of Figures

1.1	Concept of SARA: a software system for the security and safety community to share information about the AIS-related anomalies they uncover to improve maritime situational awareness.	1
3.1	Taxonomy of AIS-related anomalies, where only the first leaves are displayed. Level 1 anomalies are on the left side while level 2 is on the right side.	16
3.2	Taxonomy of AIS-related anomalies: details of the Static fields for Error in AIS message.	18
3.3	Taxonomy of AIS-related anomalies: details of the Voyage fields for Error in AIS message.	19
3.4	Taxonomy of AIS-related anomalies: details of the Navigation fields for Error in AIS message.	20
3.5	Taxonomy of AIS-related anomalies: details of the level 2 anomalies.	21
3.6	Misused VHF calls taxonomy.	25
3.7	Kinematics anomaly taxonomy [35].	26
4.1	Scope of the NIEM.	28
4.2	NIEM Information Exchange Models.	30
4.3	Anomaly Exchange XML Schema Definition (XSD).	35
4.4	Anomaly report request parameters.	36
4.5	Record metadata element.	37
4.6	Security attributes element.	38
4.7	Suspicious Activity Report design.	44

6.1	A high level overview of SARA's architecture. 1. Users are expected to interact with SARA through a trusted system such as a web application or other ECDIS. This client system provides the human interface (the UI) required to help the user build, send and receive queries to and from SARA. 2. The application server processes requests from users through the service interface and maps them to the database. 3. The database stores anomalies, ship profiles, user profiles and metadata. 4. The data processing unit is in charge of calculations and algorithms required to support the database. These include the calculation of reliability indexes for ships, creation of ship profiles and other routine maintenance processes.	54
6.2	Inside the application server. 1. All communication with clients goes through the service interface either through REST or SOAP messages. 2. The authentication service allocates each user a timed token after a successful login. Every subsequent query needs to also send the token in order for the request to be valid. 3. The core purpose of the application server is to map incoming anomaly requests into database queries. These queries include both submissions of new anomalies as well as requests for stored anomalies. 4. Operators may also edit ship profiles through queries or request to look at stored ship profiles. 5. The creation and management of user accounts also occurs through the service interface but requires different access rights. Only designated administrators (See Section 5.3.4) can create new accounts and edit security level access for example. Other user profile details can be edited by the account owner.	56
6.3	The best case scenario flowchart for submission of new anomaly reports. 1. Validation with error specific messages. 2. Immediate storing of anomalies not containing information identifying the suspect vessel. 3. The identification of a ship in SARA's database from anomaly report information. Note that this is a complex process and risky in terms of likelihood of success, since there may be many matching ship candidates with no clear singular result, and in terms of performance since the search procedure may be very computationally expensive. 4. The creation of a new ship profile when a matching ship profile was not found in the database. As this process is automated, there is a risk it will create duplicate ship profiles from reports containing spelling errors or other input discrepancies. 5. This process links an anomaly report to a ship profile. 6. The anomaly is stored in the database. 7. A successful transaction message is returned to the operator.	58
6.4	The flowchart for retrieval requests of stored anomaly reports. 1. Input validation ensures the request fields respect requirements. 2. The query is performed by filtering on record metadata with the appropriate security levels. 3. The result set is formatted and sent to the user.	60
6.5	Ship profile editing flowchart 1. Validate that the query is issued by a ruling authority account and vessel information fields. 2. Search for the ship in the database. 3. Verify that the ruling authority has sufficient security access to edit the desired records. 4. Edit the ship profile and store the changes in the database. 5. Return a successful transaction message.	62

6.6	Ship profile retrieval flowchart. 1. Validate vessel information fields. 2. Search for the ship in the database. 3. Verify that the operator has sufficient security access to retrieve the desired records. 4. Retrieve the ship profile and return the result to the user.	63
6.7	User creation flowchart. 1. SARA validates the query was sent by an administrator. 2. The new user information is validated for correctness and against potential existing accounts in SARA's user database. 3. The new user account is created and stored in the database. 4. The user handle and temporary password are sent to the user.	65
6.8	Service architecture in the context of NIEM. SARA acts as a service provider to a 3rd party application (the consumer system) through which the users (operators) interact. Both the consumer and provider maintain a copy of the trust fabric, a certificate issued by a NIEM certificate authority listing all members vetted for the network. Additionally, all communication between consumer and provider must be preceded by a SAML assertion, signed by the consumer.	66
6.9	Sequence diagram of communication exchange. All communication between consumers and providers takes place over Hypertext Transfer Protocol (HTTP) with a Secure Socket Layer (SSL) which ensures the encryption of messages. User login information is transmitted over a SAML assertion signed by the service consumer, therefore ensuring the message's authenticity. The service provider creates user sessions if all the security information is validated and service messages can continue without any additional overhead. The service consumer can POST to a logout Uniform Resource Identifier (URI), as a courtesy to SARA, to terminate the user session and free up resources.	68
6.10	Unified Modeling Language (UML) Design of the Vessel Profile data structure. . .	70
6.11	NIEM-M design of the vessel type.	71
6.12	NIEM-M design of the vessel augmentation type.	72
6.13	UML Design of the Vessel data structure.	73
6.14	NIEM-M design of the anomaly.	74
6.15	UML Design of the Anomaly data structure.	74
6.16	NIEM-M design of the contact information data structure.	75
6.17	NIEM-M design of the organization data structure.	75
6.18	NIEM-M design of the person data structure.	75
6.19	NIEM-M design of the person name data structure.	76
6.20	UML Design of the Contact Information data structure.	77
6.21	NIEM-M design of the block metadata.	78

6.22 Database design in UML.	79
6.23 SQL Server versus MongoDB representation of an XML document.	83
6.24 The Data Processing unit. 1. Since incoming anomalies may contain only partial information regarding the identification of a ship, establishing the relation between an anomaly and the vessel it references may not be straightforward. This process is therefore in charge of linking anomaly reports to the ship profiles internal to SARA. 2. A ship's reliability rating is a number that represents the accumulated product of all of the anomaly reports referencing the vessel. The rating is a representation of the severity of a ship's transgressions through its anomaly reports. The decoupling of the reliability rating from the application server is particularly useful when anomalies are being reported by an automated process. 3. Anomalies can be generated autonomously by a unit processing messages stored in an AIS database. 4. Level 1 anomalies are generated by automated processes analysing AIS reports. Level. 2 anomalies represent a deeper level of meaning requiring more processing power than the application server can afford to liberate.	84
6.25 Subcomponents of vessel identification. A fuzzy search algorithm looks through existing vessel names, MMSI and IMO for similarities with a reported vessel. A local cache of data may mirror the vessel identification data stored in SARA's database, thus freeing up SARA to handle operator requests while the identification algorithm keeps working. A list of possible matching candidates is extracted and ordered by likelihood.	85
6.26 Maritime Information Sharing Environment (MISE): All member systems communicate over the internet with the ISI which provides services for acquisition and dissemination of data between trusted systems. All systems must communicate through the NIEM exchange model and specify the attributes for access control. .	89

List of Tables

3.1	Selected information quality dimensions describing level 2 anomalies.	22
3.2	Level 2 anomaly categories.	24
4.1	Description of the vessel identification data elements.	39
4.2	Description of the anomaly identification data elements.	40
4.3	Description of the record metadata elements.	43
6.1	Overview of SQL and NoSQL differences, from [56].	82
A.1	Description of the ship anomaly upload use case.	A-3
A.2	Description of the request ship anomaly use case.	A-4
A.3	Description of the request ship profile use case.	A-6
A.4	Description of the ship profile edition use case.	A-7
A.5	Description of the user profile creation use case.	A-8
A.6	Description of the user profile edition use case.	A-9

This page is intentionally left blank.

ACAN	Advance Contract Award Notice
ACID	Atomicity, Consistency, Isolation, Durability
AIS	Automatic Identification System
API	Application Programming Interface
C2	Command & Control
C4I	Command, Control, Communications, Computers and Intelligence
CA	Certificate Authority
COE	Common Operating Environment
COG	Course Over Ground
CORA	Centre for Operational Research and Analysis
COTS	Commercial Off-The-Shelf
CPU	Central Processing Unit
CSD	Coalition Shared Data Server
DISA	Defence Information System Agency
DND	Department of National Defence
DoD	Department of Defence
DRDC	Defence Research and Development Canada
ECDIS	Electronic Chart Display and Information System
EIEM	Enterprise Information Exchange Model
ELINT	Electronic Intelligence
ETA	Estimated Time of Arrival
GCCS-M	Global Command and Control System - Maritime
GML	Geography Markup Language
HFSWR	High Frequency Surface Wave Radar
HPACCS	Halifax Port Authority Command and Control System
HTTP	Hypertext Transfer Protocol
IC	Intelligence Community
IC-ISM	Intelligence Community Metadata Standard for Information Security Marking

IEPD	Information Exchange Package Documentation
IMIC3	Interdepartmental Maritime Integrated Command, Control and Communications
IMO	International Maritime Organization
IQ	Information Quality
ISI	Information Sharing Infrastructure
JSON	JavaScript Object Notation
LRIT	Long Range Identification and Tracking
MCDV	Maritime Coastal Defence Vessel Project
MDA	Maritime Domain Awareness
MID	Maritime Identification Digit
MISE	Maritime Information Sharing Environment
MMSI	Maritime Mobile Service Identity
MSOC	Marine Security Operations Centres
MSARI	Maritime Situational Awareness Research Infrastructure
MSSIS	Maritime Safety and Security Information System
NIEM	National Information Exchange Model
NIEM-M	National Information Exchange Model - Maritime
PAL	Provincial Aerospace
RCMP	Royal Canadian Mounted Police
RD	Research & Development
RDBMS	Relational Database Management System
REST	Representational State Transfer
RJOC	Regional Joint Operations Centres
ROT	Rate Of Turn
S-AIS	Satellite-Automatic Identification System
SAML	Security Assertion Markup Language
SAR	Suspicious Activity Report
SARA	Sharing AIS-Related Anomalies

SCONUM	Ship Control Number
SHINNADS	SHip's INtegrated Navigation And Display System
SODA	Service Oriented Defense Architecture
SOG	Speed Over Ground
SSL	Secure Socket Layer
UI	User Interface
UID	Unique Identification Number
UML	Unified Modeling Language
URI	Uniform Resource Identifier
URL	Uniform Resource Locator
VOI	Vessel Of Interest
VHF	Very High Frequency
VSS	Vessel Selection System
XML	Extensible Markup Language
XSD	XML Schema Definition

This page is intentionally left blank.

Part 1

Introduction

The general concept of SARA is illustrated at 1.1. With this technology, anomalies are discovered in Automatic Identification System (AIS) data, either by operators or analysts from the maritime security community or by systems automatically sifting through AIS databases. Submitting users can be Navy or Coast Guard officers, port authority security operators or commercial flight operators. These anomalies are reported and made available to the maritime security community. The end-users are typically working in a Command & Control (C2) environment, like the one used by navy operators at sea or by the Marine Security Operations Centres (MSOC). Since the different parties involved with SARA are already using AIS-related systems, SARA is designed to be integrated with different third party systems, either to report anomalies, to detect anomalies in AIS data or to display AIS data on a map.

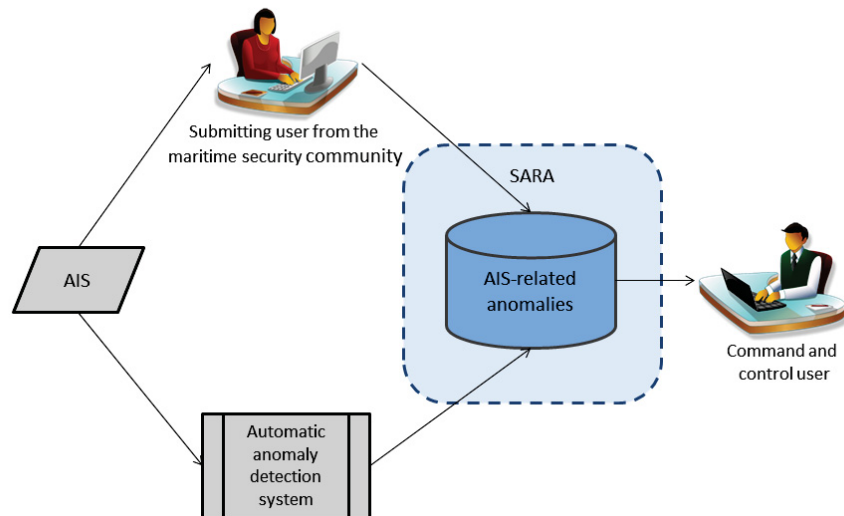


Figure 1.1: Concept of SARA: a software system for the security and safety community to share information about the AIS-related anomalies they uncover to improve maritime situational awareness.

SARA should go beyond merely indicating which ships have a history of suspicious incidents, it should indicate what sort of anomalies these were and even provide some indication of the evidence. By doing so, the main benefits of SARA would be to:

- Persist and exploit operators experience about anomalies. Such knowledge is usually lost between operations and different communities of practice.
- Provide access to information difficult or impossible to get without SARA. Anomalies observed by humans are usually not shared.
- Draw attention to vessels that might have gone unnoticed without SARA.
- Better exploit AIS data by extracting actionable information.
- Provide sufficient evidences to identify vessels of interest and allow users to drill down into anomaly metadata.

This document is organized as follows :

- Section 2 summarizes an investigation on possible collaborations opportunities for SARA with organizations selling products for the display and storage of AIS data.
- Section 3 presents a taxonomy of AIS-related anomaly types, allowing their classification and easing the sharing of prior experience.
- Section 4 examines how to represent the metadata and information elements of an anomaly for efficient discovery and sharing, with a focus on National Information Exchange Model (NIEM).
- Section 5 lists the requirements for SARA.
- Section 6 presents the software design concepts behind SARA.
- Section 7 presents the main risks associated with the development of SARA.
- Section 8 presents the application produced to illustrate the value of SARA.
- Section 9 summarizes the information learned during the realization of this call-up and serves as the general conclusion of this document.

The document also includes two annexes:

- Appendix A presents six use cases for SARA.
- Appendix B describes the demonstration application installation and launch instructions as well as troubleshooting.

Part 2

Investigation of Collaboration Opportunities

This section summarizes an investigation on possible collaborations opportunities with organizations selling products for the display and storage of AIS data.

SARA would greatly benefit from collaboration between respected players in the maritime security sector. Such a collaboration would also minimize duplication of efforts among members of the community members and ease the potential client's adoption of SARA. If SARA is integrated to software that is already part of the end user's surveillance process, the impact of adopting it will be lessened and its added value would be easier to advertise. Lastly, the feedback loop formed with clients is essential to SARA's advancement. The more SARA is exposed, the better and faster it can be adapted to end-user needs and thus be adopted by them.

This section is organised as follows :

- Section 2.1 describes the strategy used to identify potential collaborators.
- Section 2.2 lists the main Canadian maritime security and safety community actors.
- Section 2.3 describes the main products for the display and storage of AIS data that SARA's potential end users are already using.
- Section 2.4 provides a list of potential collaborators based on the product's analysis.

2.1 Methodology

Potential collaborators are organizations selling products for the display and storage of AIS data for Canadian marine security actors.

The first step was thus to list those marine security actors and then investigate the systems they are using. Systems used for the display, and storage of AIS data are not straightforward to find. Web

searches using key words like *AIS* and *Electronic Chart Display and Information System (ECDIS)* along with each of the listed Canadian marine security actors were conducted.

Once systems have been listed, research was conducted to learn if any work has been done about anomalies and if an integration of SARA would be technically possible with that system. Information was gathered from the web or directly from providers when possible.

2.2 Canadian Marine Security Main Actors

Many federal departments and agencies engage in marine security activities in their day-to-day operations:

- Department of National Defence,
- Canadian Coast Guard including Marine Communication and Traffic Services operators,
- Transport Canada with Canada Port Authorities,
- Canada Border Services Agency,
- Great Lakes and St. Lawrence Seaway system; and,
- Royal Canadian Mounted Police (RCMP).

Ideal SARA end-users operate in a **command and control environment**, with a focus on security. They have access to sources of information other than AIS, e.g. visual contacts, radars, ship manifests, etc.

From that perspective, that makes the RCMP less interesting. They usually receive tips from watch officers and act on those. They are typically not concerned with building a maritime picture and are not well positioned to identify anomalies.

2.3 AIS Display and Storage Main Products Used by the Canadian Community

This section describes the main products for AIS display and/or storage used by the Canadian marine security actors named in section 2.2.

Potential collaborators are the organizations selling and/or building these products. Some products have been discarded from this analysis because of the limited use within the Canadian community.

Each subsection is divided as follows:

1. product's description,

2. clients,
3. integration with SARA* and
4. anomaly work*.

*Note that information about integration and anomaly work is not provided for all products. The information openly available on the web has been gathered here but in some cases none was available. All other information provided during private conversations has been left out due to its potentially sensitive nature.

2.3.1 ECPINS/SHINNADS

ECPINS is a computerized, shipboard navigational aid that displays electronic charts, own vessel's position in real time, and sensor data [1].

ECPINS, also called SHIP's INtegrated Navigation And Display System (SHINNADS) for the customized Royal Navy version, is developed and maintained by OSI Maritime Systems [2]. Installation across the Royal Canadian Navy surface and subsurface fleet started in 1989. The system will continue to be maintained and upgraded until 2020, with options to extend for up to an additional 20 years.

Warship-AIS (W-AIS) is a module part of ECPINS integrating IMO-compliant AIS. This capability is an ECPINS spin-off resulting from work with the United Kingdom Royal Navy. The Warship functionality allows users to manipulate data through database analysis, create contacts and share information between other W-AIS users [3]. The system enables the tactical exploitation of commercial AIS and other information collected by the sensors integrated with ECPINS. W-AIS is now part of SHINNADS and thus deployed fleet-wide in the Royal Navy.

2.3.1.1 Clients

ECPINS is installed across the Royal Canadian Navy surface and subsurface fleet and on at least 12 coastal vessels of the Canadian Coast Guard fleet [4].

2.3.1.2 Integration

SHINNADS has a flexible open interface and seamlessly integrates with third party products. So it is safe to assume that it would be technically possible to integrate SARA to SHINNADS.

However, user interface modification is not straightforward, as there are multiple constraints imposed by the Royal Navy about how information should be displayed on a screen on board of a ship.

2.3.2 Interdepartmental Maritime Integrated Command, Control and Communications (IMIC3)

IMIC3 is a data exchange system that allows Royal Canadian Navy and Coast Guard vessels to securely exchange information in near real time, and collaborate on a common maritime picture, in order to improve decision making. Vessels equipped with IMIC3 send back data from their local operating area to the MSOCs, which then integrate this information into a larger picture that can be transmitted to all ships [5]. Sensing capabilities includes AIS and fitted sensors/sources that may be unique to each vessel, such as automated radar plotting systems. This information is fused with Long Range Identification and Tracking (LRIT) and geospatial areas of activities to provide situational awareness.

The project was awarded to Thales Canada in March 2011 and delivered in October 2014. It uses Thales Canada's COMMANDER C3 non-combatant solution (see [6] for details).

Thales Canada is also developing seven portable Commander C3 systems. Based on a laptop and a SATCOM terminal, these could be set up quickly on smaller vessels, such as RCMP or Department of Fisheries and Oceans craft [7].

2.3.2.1 Clients

IMIC3 is installed on board of:

- 12 Royal Canadian Navy Kingston class coastal vessels and
- up to 44 sea-going Canadian Coast Guard vessels.

2.3.2.2 Integration

Commander C3 brochure [8] mentions that the system provides flexibility and simple integration with existing systems and equipment. It also provides an efficient protocol for reliable communication and data exchange between Commander C3 nodes in the system. Therefore, it is safe to assume that it would be technically possible to integrate SARA to Commander C3.

2.3.3 GCCS-M

Global Command and Control System - Maritime (GCCS-M) is the Royal Navy's primary fielded command and control system installed on board of frigates and in MSOCs.

GCCS-M is the maritime implementation of the GCCS family of systems. It supports decision making at all echelons of command with a single, integrated, scalable Command, Control, Communications, Computers and Intelligence (C4I) system. As described in [9], the C4I system fuses, correlates, filters, maintains, and displays location and attribute information on friendly, hostile,

and neutral land, sea, and air forces, integrated with available intelligence and environmental information. It operates in near real-time and constantly updates unit positions and other situational-awareness data. GCCS-M also records data in databases and maintains a history of changes to those records.

As GCCS-M processes multiple data sources, including Electronic Intelligence (ELINT), Provincial Aerospace (PAL) data and High Frequency Surface Wave Radar (HFSWR), as well as AIS data.

GCCS-M is supported by the Common Operating Environment (COE) architecture. The COE is an environment for collaborative software development and execution developed by the US Defence Information System Agency (DISA). Technically, GCCS-M is a system made of a group of COE segments.

It is important to note though that there are ongoing investigations to help identify the requirements for the future GCCS-M replacement. However, the replacement date is not determined yet.

2.3.3.1 Clients

GCCS-M is installed on board of Canadian Royal navy frigates and in MSOCs.

2.3.3.2 Integration

GCCS-M can only manage a limited number of tracks. This characteristic, added to the fact that GCCS-M will be soon replaced, greatly limit SARA's integration potential to the actual GCCS-M. If integration is still an option, SARA would have to be developed as a COE segment.

In the future version of GCCS-M, SARA would be probably a web service as the most probable future framework would be cloud-based. In the meantime, SARA could be developed as a web service and integrated in the DRDC-Ottawa Service Oriented Defense Architecture (SODA) environment. However, SODA is for Research & Development (RD) purposes, not in an operational context.

2.3.4 Halifax Port Authority Command and Control System

The Halifax Port Authority Command and Control System (HPACCS) displays radar tracks on electronic charts, annotates them with data transmitted by the vessel's AIS transponder and automatically cues cameras onto targets of interest using intelligent object recognition, tracking and scene analysis [10].

This project involved more than 20 subcontractors and vendors. The command and control system was designed by Ultra Electronics and integrates the PureActiv camera system from PureTech. The Ultra solution provides information and secure access to regional users such as port police, port operations staff, first responders and harbour pilots. Users are able to access the system through a web portal and view real-time vessel locations incorporating AIS data and Lloyd's

registry information. It also provides a comprehensive port management suite: vessel scheduling and berth management, harbour pilot scheduling, billing and maritime security alert management [11].

The contract was awarded to Ultra in April 2007 and the installation was completed by June 2012 at the Halifax Port Offices in downtown Halifax.

2.3.4.1 Clients

HPACCS was developed for the Halifax Port Authority.

2.3.5 Provincial Aerospace

PAL provides intelligence, surveillance and reconnaissance and maritime patrol aircraft operations and systems. They do planning and execution of mission operations including the collection, analysis and dissemination of data [12].

Surveillance aircraft have fully integrated AIS receiver on board. The AIS permits the aircraft to collect information that is automatically transmitted from ships from a distance of up to 200 nautical miles.

2.3.5.1 Clients

- Transport Canada,
- Fisheries and Oceans Canada,
- Environment Canada,
- Department of National Defence (search and rescue operations) Victoria JRCC,
- Canadian Coast Guard and
- possibly other federal agencies.

2.3.5.2 Integration

Because PAL aircraft are equipped with multiple other data sources, operators on board are in a good position to detect and report AIS data anomalies. Therefore, PAL would be a great AIS-related anomaly provider for SARA. Moreover, PAL uses its own system to report data, which opens doors to integration.

2.3.6 TimeCaster

TimeCaster is a Maerospace product that fuses satellite AIS, coastal AIS and other data sources. It uses this fused data to predict the current and future positions of all the AIS-equipped ships in the world. This prediction uses patent pending algorithms [13] going beyond dead reckoning of selected ships [14].

2.3.6.1 Clients

The target clients are navies, coast guard and rescue, customs and immigration departments, fisheries and environment ministries, port authorities, shipping industry and commodities trading firms.

In June 2015, Maerospace obtained a contract with Canada's Department of National Defence (DND) to use and evaluate TimeCaster service for one year in support of DND global maritime domain awareness mission. With the finalization of the DND contract, the TimeCaster service implemented live operations in secure DND Monitor facilities [15]. This contract also gives access for one year to any authorized government member.

2.3.6.2 Integration

TimeCaster is built on the Maerospace scalable information service-based platform. This web-based architecture is flexible making it is possible to interact with TimeCaster components [16]. Therefore, TimeCaster could be used as an automated anomaly reporter for SARA.

2.3.6.3 Anomalies

TimeCaster does anomaly detection and can send predictive alerts identifying anomalies before they are reported by the ship. The system is adapted to track, forecast, and detect anomalies in a vessel's reported and predicted positions. The operator can define geographical and attribute filters and alerts (email, pop ups, ...). These alerts will be triggered each time an anomaly related to ships corresponding to the filter criteria is detected and/or predicted up to four hours in the future.

TimeCaster is adapted to evolve a list of anomalies associated with AIS messages and ship behaviour. These anomalies fit one of two classes: intrinsic and behavioural. Anomalies tagged and reported by TimeCaster can be found in the patent claim (see [13]) and can be augmented.

2.3.7 exactEarth

exactEarth is a Canadian company delivering Satellite-Automatic Identification System (S-AIS). They offer a range of products all based on the S-AIS data they gather:

- exactAIS feed and archives (back to 2010).
- ShipView: web-based viewing tool that allows users to see all the ship positions produced by the exactAIS data service and then plots them on a set of map layers to enhance the viewing experience.
- exactAIS Density Maps: vessels pattern analysis.
- exactAIS Trax service: small vessels tracking with AIS Class B type receivers.
- Information services: value-added Information-as-a-Service offerings based on big data analytics and processing of the raw AIS data being received through the exactView processing chain.

2.3.7.1 Clients

The data is available for all Government of Canada agencies and departments, including the MSOCs, and has been used to support security and surveillance, environmental and fisheries protection and enforcement, Arctic monitoring and search and rescue events. In 2014, the Canadian government renewed with a two-year contract for exactAIS (including archives).

2.3.7.2 Integration

An integration would not be possible with the exactAIS product, since it is a data feed without any decision support system.

ShipView could be a potential product for integration with SARA. However were unable to find out if any of the Canadian marine security actors listed in section 2.2 is using it.

Another option could be to integrate their anomaly-related analysis products (see next section) as a source of anomalies for SARA, but the technical feasibility of that option was not confirmed.

2.3.7.3 Anomalies

exactEarth has started different initiatives on anomaly detection and consistency checking within AIS data. They are described in the preliminary prospectus issued in June 2015 for initial public offering of its common shares (descriptions below are excerpts from the prospectus [17]).

exactEarth has marketed a Positional Anomaly Services that detects and reports on suspicious and/or erroneous AIS location messages, based on processing of the true behaviour and movements of the vessels regardless of the abnormal information being transmitted. They produce daily reports for their customers on such behaviours and anomalies on a subscription basis.

Another information product is the Knowledge Attributes, which is information derived from new and interesting aspects of the ship behaviour. Ship rendezvous, ships stopping in open ocean, and

ships deviating from shipping lanes are all examples of value added information in the form of Knowledge Attributes that are being made available, on a subscription basis, to customers.

exactEarth also provides Voyage History and other Behavioural Reports. Such reports range from providing information on all ships heading towards a particular location, along with estimated arrival times, to ships transiting or encroaching a specific area of ocean, or even to historical look-backs to the ports visited by a selected ship over a period of time, with additional information on the amount of time spent in each port.

exactEarth also filed a patent in 2012 called Method for Consistency Checking and Anomaly Detection in AIS Data [18]. These methods aim at detecting spoofed AIS messages. More precisely, they aim at validating positions reported in the AIS message with signal characteristic data, such as timing and Doppler shift data, derived from a plurality of AIS message signals. Vessels whose reported positions deviate from the fitted function may be flagged as suspect.

2.3.8 Vessel Selection System

The Vessel Selection System (VSS), developed by Greenline Systems, an A.T. Solutions company, helps analysts and decision-makers identify targets of interest and make interdiction decisions. VSS develops a maritime picture by integrating data from multiple sources, such as terrestrial and satellite based ship position AIS data, vessels and company characteristic data; entity or vessel watch lists; and agency databases containing details on the vessel's cargo and crew. It allows users to visualize current and past vessel tracks against IHS Fairplay database information. These data are analyzed and presented with an interface supporting a collaborative work environment [19].

2.3.8.1 Clients

VSS is currently used by the MSOCs to display real-time and historical vessel data, in addition to other references and operational data layers from MSOC partners. It allows for the identification, assessment and reporting of maritime activities that represent a potential threat to the sovereignty, security and safety of Canada. VSS is going to be used by MSOCs until at least 2019. See details in the 2013 Advance Contract Award Notice (ACAN)[20].

2.3.8.2 Integration

As mentioned in the 2013 ACAN, the MSOC Greenline installation had to be integrated with a number of other Commercial Off-The-Shelf (COTS) products. As a result, the Greenline tool suit had to be customized to allow it to work within the MSOC infrastructure, to accept additional data feeds from some of the MSOC partners, and to seamlessly integrate with other MSOC sub systems and products including Geo-Spatial Information Systems, Maritime Interest Management systems, and the MSOC Information Portal.

Moreover, it is possible to extend VSS with additional datasets without any customization [21].

It can thus be assumed that VSS could be customized to include SARA functionalities. Moreover, the option of adding vessel's anomaly history as a data source could be a simple option that should be investigated.

2.3.8.3 Anomalies

VSS includes a default rule set enabling users to analyze risks. Each rule is associated with a risk score and that score is applied when a rule is activated against a vessel or entity. The rules engine awards a cumulative score, which represents the risk level for that vessel or entity. The rules engine is highly customizable, so an organization can modify and add rules and scores to address its specific areas of concern[20].

A relation can be established between the concepts of *risk* and *anomaly*. The higher the number of anomalies associated to a vessel, the higher its risk level will be. Following that concept, assuming SARA's vessel' anomaly histories can be added as a data source, it could be used in the risk assessment algorithm. For instance, a vessel with a long history of wrong or missing destinations would be considered and displayed as riskier by VSS. Different weights could also be associated to anomalies to differentiate their impact in the risk assessment algorithm. This proposition was not validated with Greenline personnel nor with MSOC operators but seems promising.

2.3.9 MSSIS-TV32/SeaVision

In 2002, Volpe - The National Transportation System, implemented a vessel communications and tracking network for the Saint Lawrence Seaway [22]. The network, based on the AIS, promises improved safety, security, and efficiency throughout the Seaway.

Volpe used its data and system engineering expertise in 2006 to develop the Maritime Safety and Security Information System (MSSIS) [23], a maritime domain awareness network, for the U.S. Navy Sixth Fleet.

MSSIS is a low-cost, unclassified, near real-time network that is used to track vessels as they traverse the world's waterways. Transview (TV32) [24], the client software for MSSIS, serves as a common system interface and vessel tracking display for its users.

Volpe also developed SeaVision, which is a web version of TV32. SeaVision has the main advantage (from this project perspective) of allowing for future expansion beyond AIS. For instance, a user could overlay radar, weather, maritime, or other user-inserted data on top of the system. In addition, SeaVision is coupled to a threat detection tool which provides alerts on vessels of interest. This tool allows the operator to define rule-based threat from AIS data.

2.3.9.1 Clients

TV32 serves as the vessel traffic monitoring system for the Saint Lawrence Seaway AIS network. But it was not possible to find whether Saint Lawrence Seaway has upgraded to this web version.

2.3.9.2 Integration

TV32 has been customized for several different sponsors: the Panama Canal, several ports in Latin America, possibly the Saint Lawrence Seaway, U.S. DoD, Columbia River Pilots and others. It also serves as a rapid prototyping tool for the U.S. Coast Guard Research and Development Centre. This indicates that TV32 can be customized, which opens the door to an integration with SARA. However, from the descriptions found, SeaVision seems to be a better candidate for integration, as it was already coupled with a threat detection tool.

2.3.9.3 Anomalies

It is not uncommon to see BRITE used with MSSIS data feed with or without TV32. Whether the Great Lakes Saint Lawrence Seaway System is using it or not could not be verified.

BRITE web services, developed by NATO, examine the stream of MSSIS to identify and resolve anomalies by checking for consistency with other data in the system (detecting duplicates, etc.) and comparing reported ship data to other reliable data sources, like the Lloyd's of London database. It provides alerts based on criteria the analyst or watch stander defines, pointing out vessels that may warrant additional attention [25]. BRITE reports kinematic anomalies but also AIS-related anomalies, such as non-existent or incorrect International Maritime Organization (IMO) number, the same IMO number being held by more than one ship, a mismatch between the IMO number and the name or call sign, etc. [26]

2.4 Potential Collaborators

The potential collaborators are listed below, from the order they appear in section 2.3. Products with less visibility or soon to be obsolete were discarded.

Potential collaborators:

- OSI Maritime Systems,
- Thales Canada,
- PAL,
- Maerospace,
- exactEarth and
- Greenline Systems.

This page is intentionally left blank.

Part 3

Taxonomy of AIS-related Anomaly Types

This section presents the taxonomy developed to represent AIS-related anomaly types. It aims to answer the questions: *What kind of anomalies show up in Automatic Identification System (AIS) data? Which of these kinds of anomalies might be of interest to end users?*

This section is organised as follows :

- Section 3.1 presents a concept of level 1 and 2 anomalies.
- Section 3.2 presents the proposed taxonomy of AIS-related anomaly types.
- Section 3.3 provides some remarks about maritime kinematics anomalies.

3.1 AIS Anomalies Levels 1 and 2

AIS Anomaly types can be defined differently depending on the point of view:

1. **AIS messages** point of view: an anomaly is any kind of error within the message, e.g. missing information, inconsistent flag, out of range position, etc. These errors can be intentional or not. These will be referred as **level 1** anomalies.
2. **Identified ship** point of view: an anomaly is a behaviour that deviates from expectations [27]. In regards to AIS messages, it implies a history of AIS messages, attached to the same uniquely identified ship, describing a deviant pattern. These are **level 2** anomalies.

3.2 Taxonomy

Figure 3.1 shows the principal branches of the proposed taxonomy of AIS-related anomalies. The root of the taxonomy is a uniquely identified ship. This ship-centric point of view is closer to what a Royal Navy or Coast Guard operator experiences in daily operations than the message-centric one.

A unique ship identification assumes, however, some sort of fusion algorithm using another source of information to remove most ambiguities about the identity. If Sharing AIS-Related Anomalies (SARA) is integrated with SHINNADS for instance, such disambiguation process is already included. If there is no integration, such mechanism will have to be part of the system. Also, a means to uniquely identify the ship will have to be developed, e.g. an Unique Identification Number (UID), Uniform Resource Identifier (URI), etc.

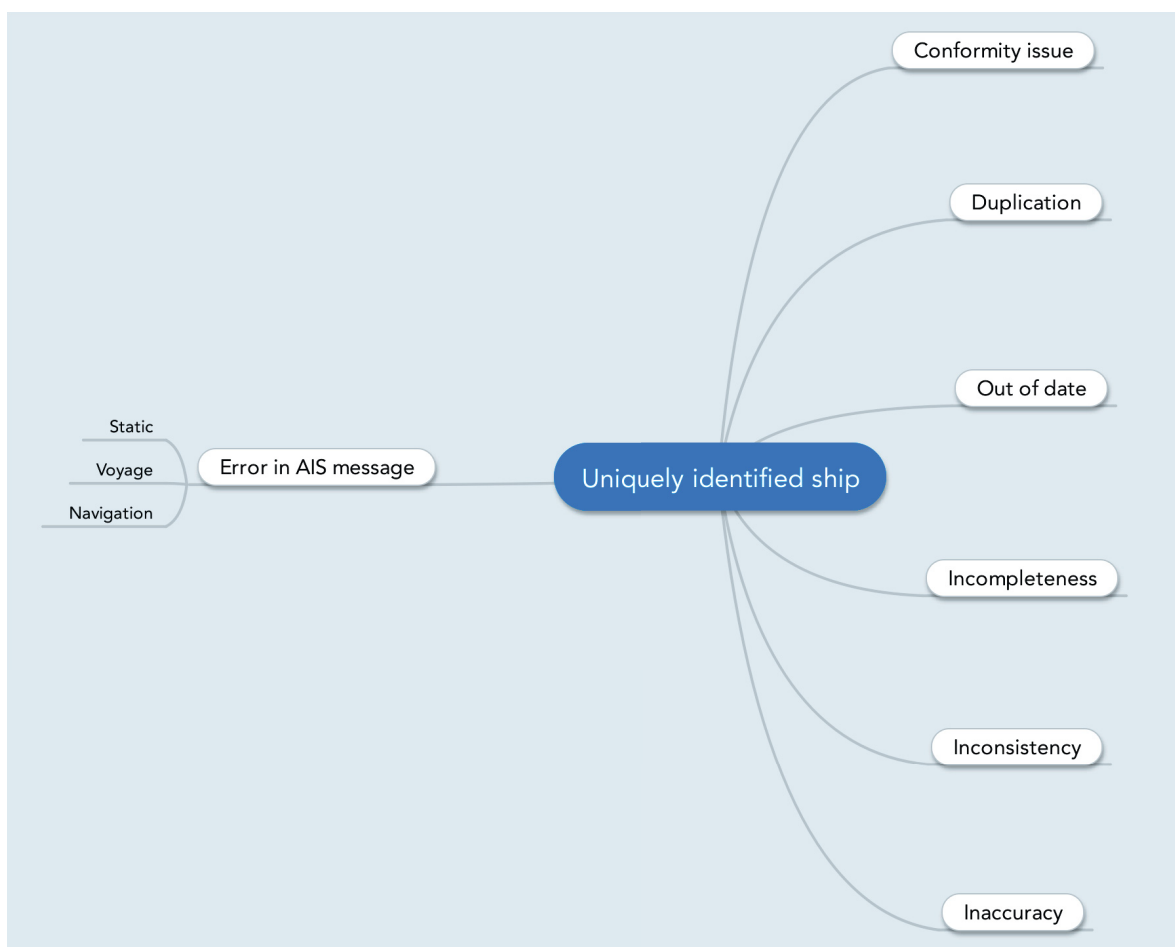


Figure 3.1: Taxonomy of AIS-related anomalies, where only the first leaves are displayed. Level 1 anomalies are on the left side while level 2 is on the right side.

3.2.1 Error in AIS Messages (Level 1)

An operator can report anomalies originating in transmissions from an identified ship. Reasons can vary, but they include the possibility that the operator does not have a full picture of the vessel behaviour but notices some anomalies within a message. Also AIS-related anomalies could be automatically reported by a system sifting through an AIS database, such as Maritime Situational Awareness Research Infrastructure (MSARI). Errors in AIS messages can be easily detected automatically, since they require no context.

The taxonomy allows for the reporting of message-centric anomalies attached to an identified ship with the Error in AIS message branch. This branch represents most of the possible errors contained in AIS message types 1, 2, 3 and 5 and is divided in 3 main sub-information types (closely related to AIS message types): static, voyage and navigational. Anomalies originating from other types of messages could be added to the taxonomy, but they are of lower priority.

1. **Static** information is entered when the AIS system is installed: Maritime Mobile Service Identity (MMSI), IMO number, call sign, name, length and beams, type of ship, etc. Figure 3.2 shows the details of the taxonomy of the error in AIS message associated with static information.
2. **Voyage** information is updated manually at the beginning each trip: draught, cargo, Estimated Time of Arrival (ETA), destination, etc. Figure 3.3 shows the details of the taxonomy of the error in AIS message for the voyage information.
3. **Navigational** information mostly derives from onboard electronic navigational system, but can also be manually modified: timestamp, Speed Over Ground (SOG), Course Over Ground (COG), Rate Of Turn (ROT), heading, etc. Figure 3.4 shows the details of the taxonomy of the error in the AIS message for the navigational information.

Errors in AIS messages are well documented in the literature. The 3 main references used to build the taxonomy are: [28], [29] and [30].

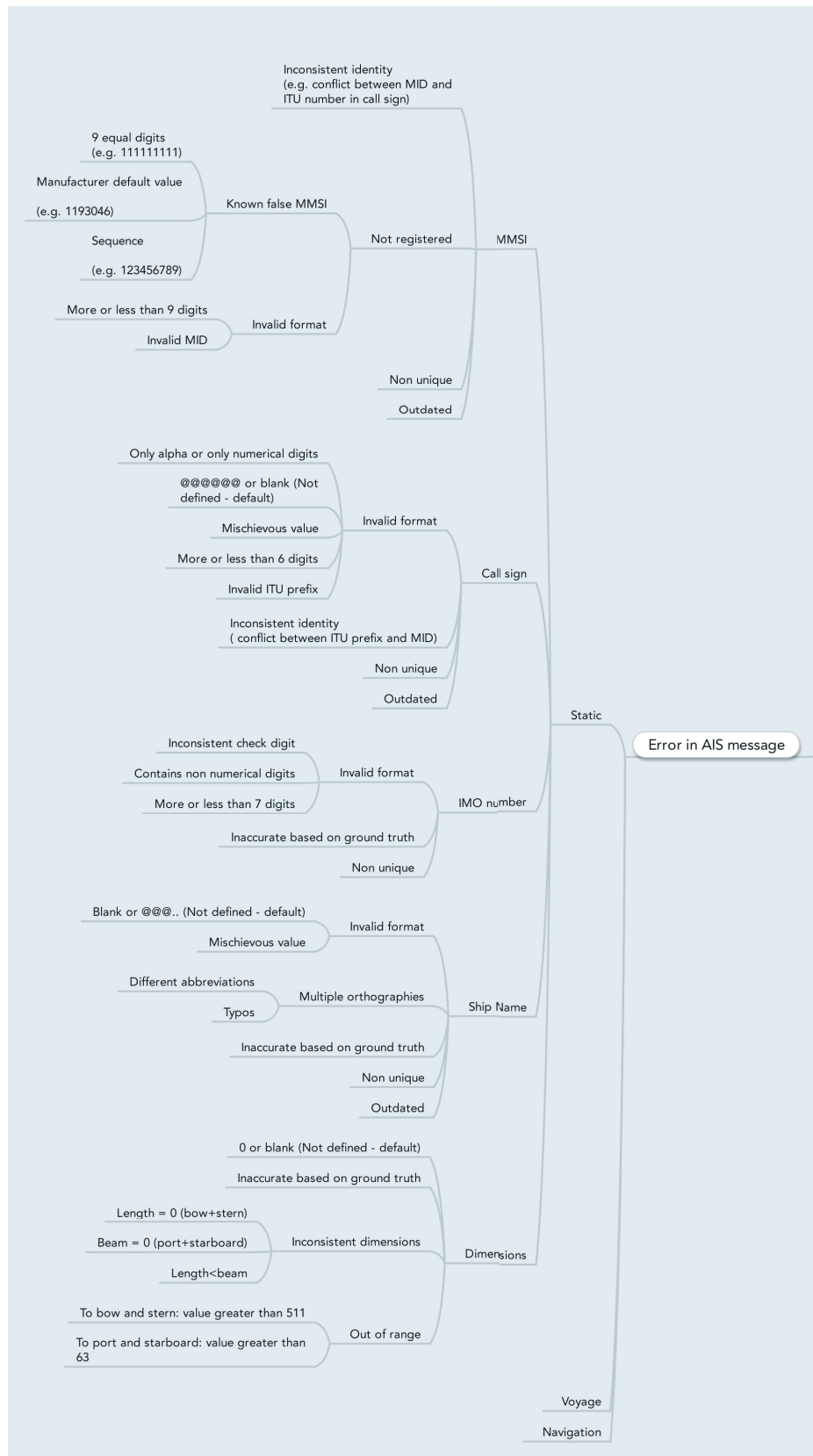


Figure 3.2: Taxonomy of AIS-related anomalies: details of the Static fields for Error in AIS message.

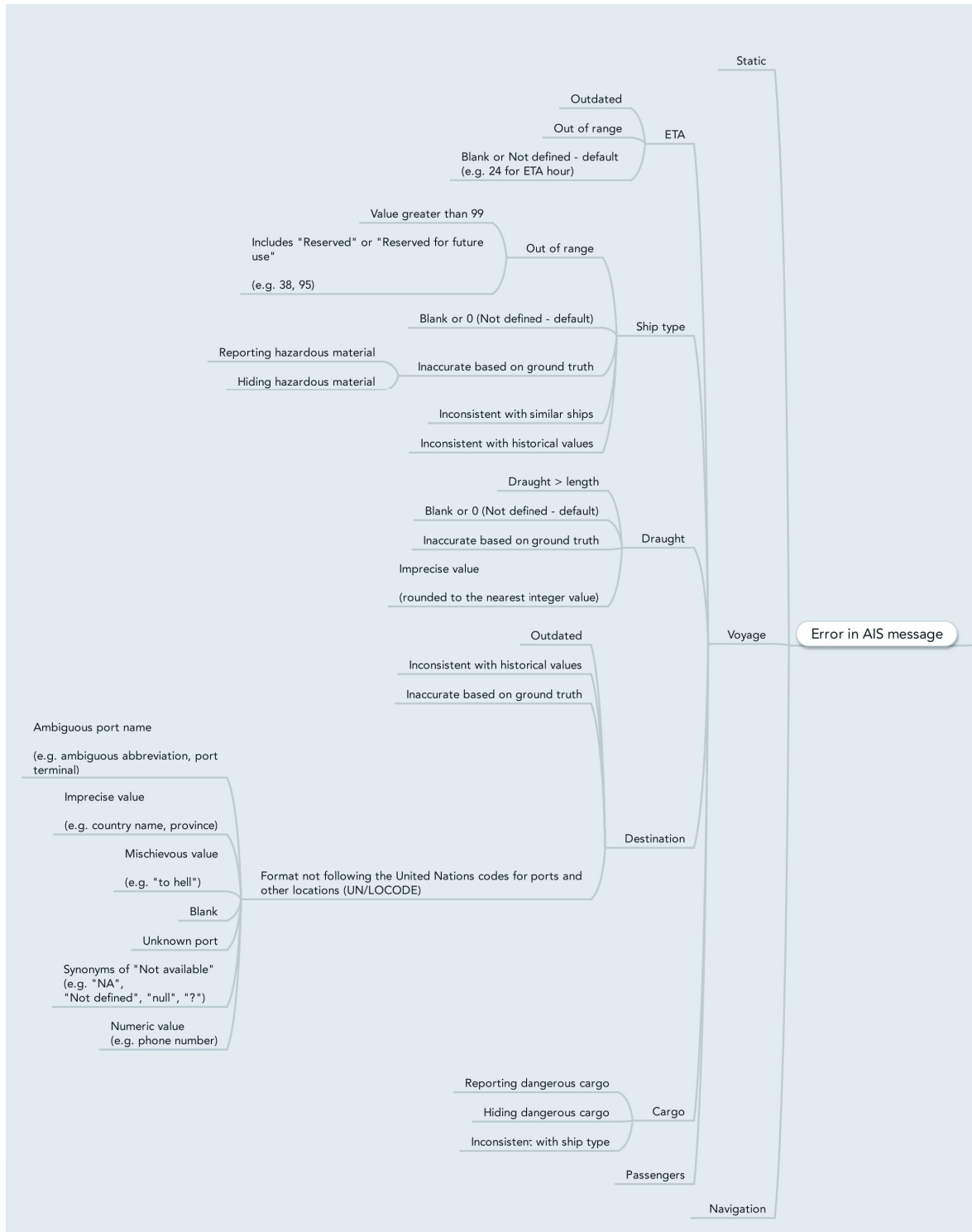


Figure 3.3: Taxonomy of AIS-related anomalies: details of the Voyage fields for Error in AIS message.

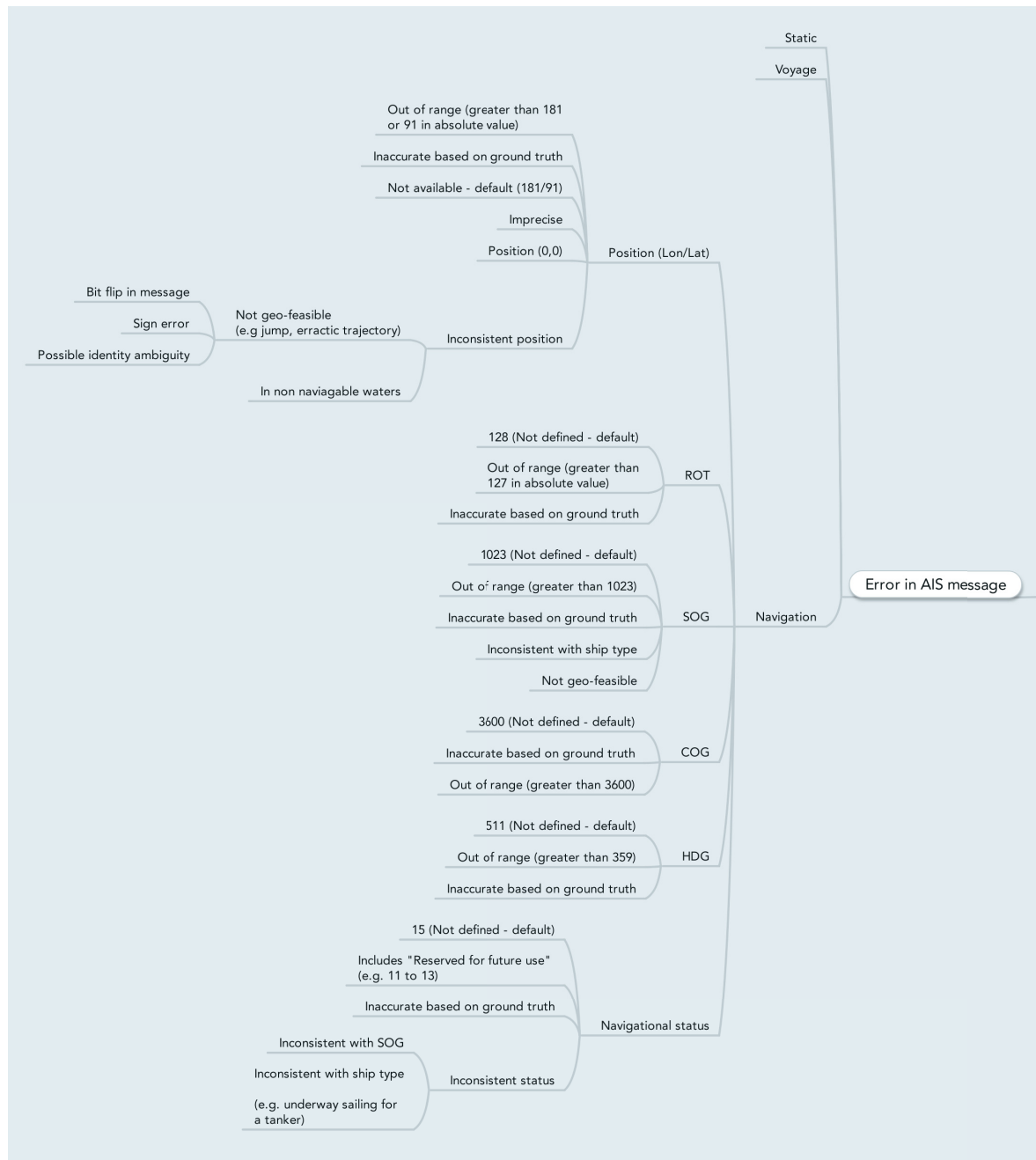


Figure 3.4: Taxonomy of AIS-related anomalies: details of the Navigation fields for Error in AIS message.

3.2.2 Transition from Anomalies Level 1 to 2

There is a link between anomalies of level 1 and of level 2. A level 2 anomaly arises from a history of AIS messages, attached to the same uniquely identified ship, describing a deviant pattern. The

right side of the taxonomy contains these kinds of anomalies. Details are illustrated in Figure 3.5.

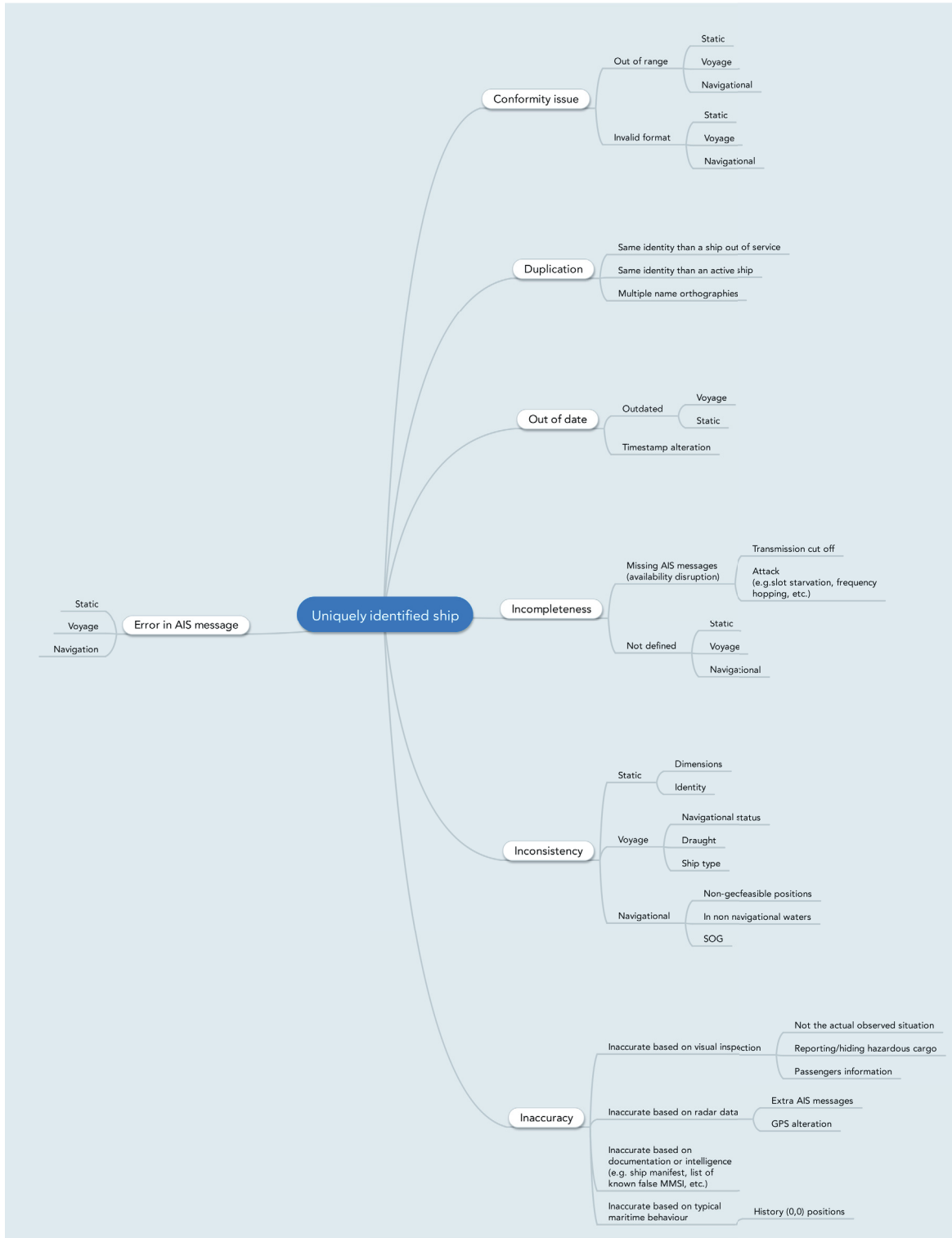


Figure 3.5: Taxonomy of AIS-related anomalies: details of the level 2 anomalies.

A single AIS message error may not be of great interest to an operator, but a history of AIS message errors can become interesting.

All level 1 to level 2 transitions had to be captured in the taxonomy to cover all possibilities. While there is more than one way to summarize level 1 anomalies, an Information Quality (IQ) approach was selected. One of the reasons to favour this approach is the option to eventually create and compute a score that would describe a ship based on its anomalies history. A similar approach was also proposed by Iphar, Napoli and Ray [31], but with slightly different IQ dimensions.

Six dimensions of IQ have been selected to describe level 2 anomalies (described in Table 3.1).

IQ dimension	Definition
Validity	Conforms to the syntax (format, type, range) of its definition.
Uniqueness	An entity is represented only once.
Timeliness	The situation-dependent degree to which information is available when needed [32].
Completeness	The degree to which the information is free of gaps [32].
Consistency	The extent to which information is in agreement with related or prior information [32].
Accuracy/correctness	The degree to which information agrees with ground truth [32].

Table 3.1: Selected information quality dimensions describing level 2 anomalies.

Based on these definitions, six anomalies of level 2 have been defined (see Table 3.2).

Anomaly type	Definition	Example	IQ dimension
Conformity issue	The ship has a history of not following the AIVDM/AIVDO format standards (see [33]).	<ul style="list-style-type: none"> History of AIS messages containing out of range positions. History of AIS messages containing a MMSI with invalid Maritime Identification Digit (MID) 	Validity
Duplication	<p>The ship has a history of:</p> <ul style="list-style-type: none"> using other ships' identity information or having other ships using its identity information. 	<ul style="list-style-type: none"> History of AIS messages containing the MMSI of an active ship. History of AIS messages with ship names spelled differently. 	Uniqueness
Out of date	The ship has a history of outdated information.	<ul style="list-style-type: none"> History of AIS messages containing outdated ETA. History of AIS messages containing the ship's former name. 	Timeliness
Incompleteness	The ship has a history of missing information.	<ul style="list-style-type: none"> Availability disruption. History of AIS messages containing non-defined ROT. 	Completeness

Anomaly type	Definition	Example	IQ dimension
Inconsistency	The ship has a history of inconsistent information. Inconsistency is detected either from conflicting information within a single AIS message or between AIS messages.	<ul style="list-style-type: none"> History of AIS messages containing navigational status conflicting with the ship type. Non-geo-feasible position. 	Consistency
Inaccuracy	The ship has a history of reporting information that is not correct. Ground truth may be: visual inspection, radar data, documentation or intelligence (e.g. ship manifest, list of known MMSI, etc.) and typical maritime behaviour (i.e. common sense).	<ul style="list-style-type: none"> A history of AIS messages containing (0,0) positions. A history of AIS messages with MMSI value 1193046 (manufacturer default value). 	Accuracy/ correctness

Table 3.2: Level 2 anomaly categories.

SARA being a reporting system, the proposed taxonomy allows for the reporting of level 2 anomalies without having to report the associated AIS message anomalies.

3.2.2.1 Time Dimension

Time is implicit to the notion of history: *When does a series of erroneous AIS messages become a level 2 anomaly? How many AIS messages with undefined navigational status do we need before flagging it to the operator? What is the time window we should consider for AIS messages history?*

We don't have all the answers at this early stage of SARA development. We believe that sort of decision should be left in the hands of the operators. They could configure the kinds of anomalies they want to monitor.

Also, part of the answers are in the type of the erroneous AIS information: static, voyage, navigation. Type 5 AIS messages (static and voyage information) are broadcast every 6 minutes. Types 1, 2, 3 (navigational information) are broadcast every 2-10 seconds while underway and every 3

minutes if stationary or at anchor. Therefore, we need at the very least twice more types 1-3 messages to cover the same period as with a type 5 message. It makes sense thus to require a longer history for navigational erroneous AIS data than for static and voyage before flagging it as a level 2 anomaly. Moreover, errors in static AIS data should be identified as level 2 anomalies quicker than voyage data because most of it was entered when the AIS system was installed and is not meant to change.

Finally, note that transitions from level 1 to level 2 anomalies could be partially done automatically with an algorithm. But the definition of the time thresholds used for automatic transitions should be carefully discussed with operators. If thresholds are too low, there will be too many anomalies triggered and SARA risks being turned off or ignored. On the other hand, if the thresholds are too high, not enough anomalies would be flagged and it may impact end-users trust in the system.

3.2.3 Low Priority Anomalies

While developing the taxonomy, the problem of misused Very High Frequency (VHF) calls (see Figure 3.6) was identified as being of low priority and discarded from the taxonomy. This work is based on [34].

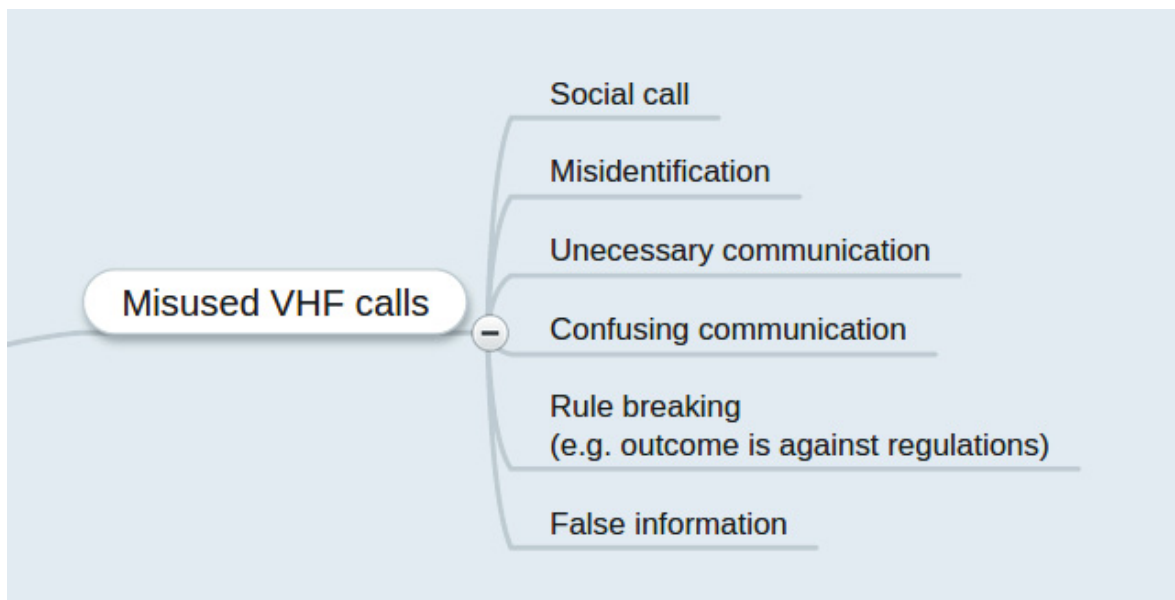


Figure 3.6: Misused VHF calls taxonomy.

3.3 Maritime Kinematic Anomalies

Operators may be interested into broadening the taxonomy to include other maritime anomalies, such as kinematic anomalies. Kinematic anomalies are out of scope at this stage of the project, but we included kinematic anomaly taxonomy developed based on the outcomes of a 2008 workshop [27]

on the topics with attendees from Defence Research and Development Canada (DRDC) Valcartier and Centre for Operational Research and Analysis (CORA), Trinity, Regional Joint Operations Centres (RJOC) Atlantic and MacDonald, Dettwiler and Associates. See Figure 3.7.

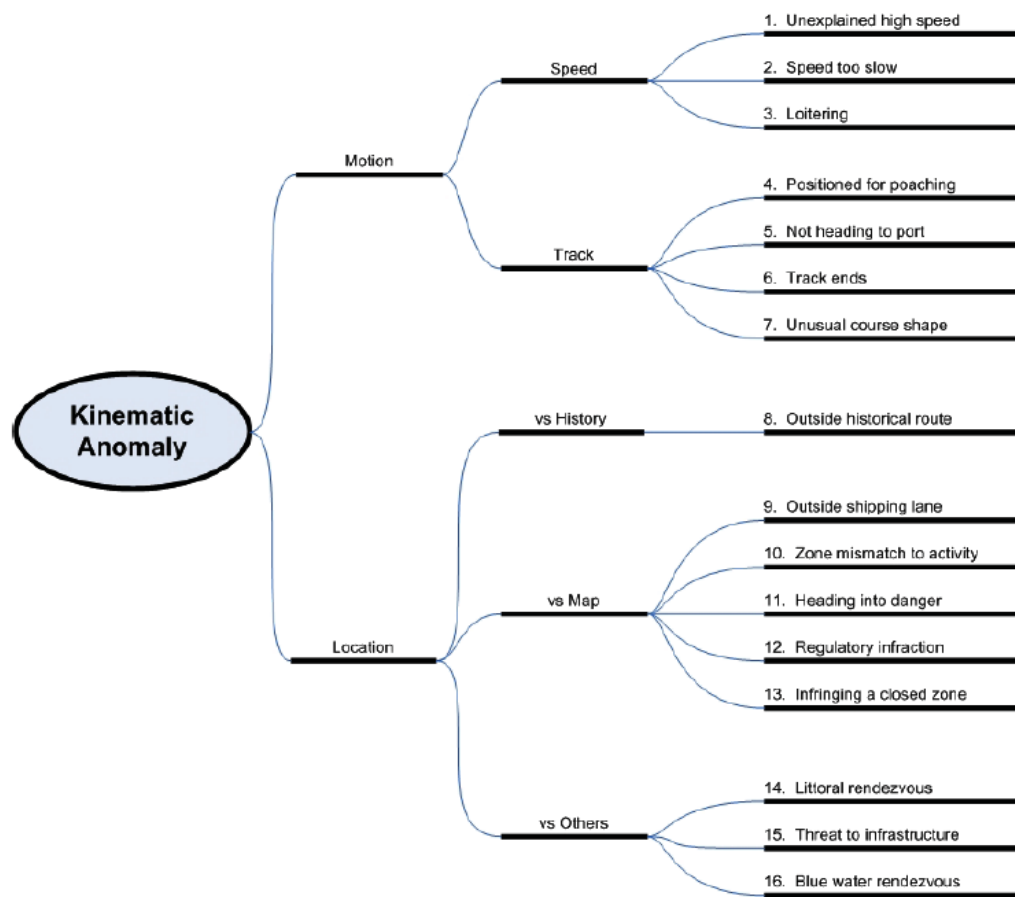


Figure 3.7: Kinematics anomaly taxonomy [35].

Part 4

Representation and Sharing of Metadata about Anomalies

This section examines how to represent the metadata and information elements of an anomaly for efficient discovery and sharing. It is based on the National Information Exchange Model (NIEM) framework, as its use is becoming more widespread within multiple domain of application including the maritime domain. This section is organised as follows :

- Section 4.1 presents an overview of the NIEM and its framework.
- Section 4.2 presents the NIEM-Maritime metadata placeholder used to describe an anomaly.
- Section 4.3 presents the security restrictions and their attributes.
- Section 4.4 proposes a NIEM-based Anomaly Exchange Report.
- Section 4.5 presents pertinent requests already defined as part of the NIEM.
- Section 4.7 presents existing Information Exchange Package Documentation (IEPD)s for sharing anomalies and lessons learned.

4.1 National Information Exchange Model

NIEM is a community-driven, standards-based approach to exchanging information. NIEM was developed in the United States and its adoption is growing in Canada [36]. Diverse communities can collectively leverage NIEM to increase efficiency and improve decision making [37].

NIEM standards enable different information systems to share and exchange information, irrespective of the particular technologies in use in those information systems. As shown in figure 4.1, NIEM is used solely as the message during data/information exchange between 2 systems.

The NIEM aims toward more efficient and expansive information sharing between agencies and jurisdictions; more cost-effective development and deployment of information systems; improved

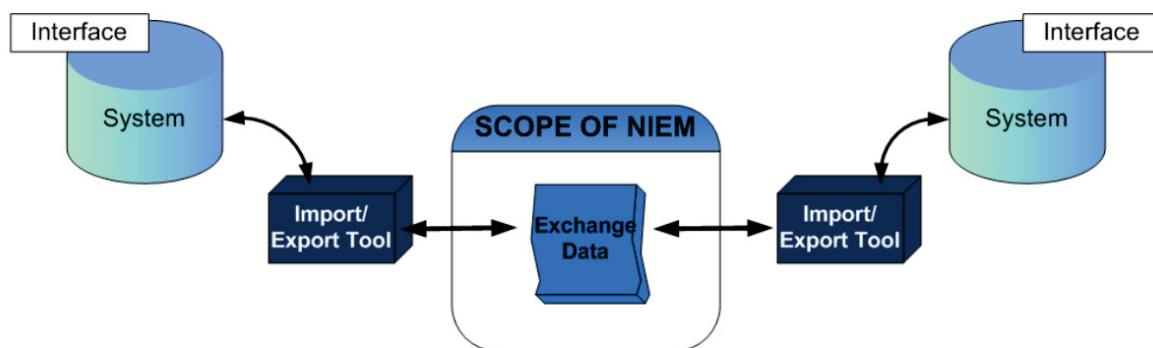


Figure 4.1: Scope of the NIEM.

operations and better quality decision making. It hopes to achieve these aims by providing more timely, accurate, and completed information; so thus enhancing public safety and homeland security. Moreover, creating and adopting NIEM standards means that multiple organizations can reap significant cost benefits through adoption and reuse, rather than building proprietary, single-use software from scratch [38].

The NIEM framework has several components:

1. NIEM Core: A common Extensible Markup Language (XML)-based data model that provides data components for describing universal objects, such as people, locations, activities, and organizations.
2. Domains: More specialized XML data models for individual use cases. There is a specialized domain for Maritime, along with a number of others (e.g. Justice and Immigration).
3. Information Exchange Package Documentation (IEPD): A methodology for using and extending the building blocks that comes from the common and domain-specific models, and turning them into a complete information exchange.
4. Tools: Help develop, validate, document, and share the information exchange packages.
5. Governance Organization: Provides training and support and oversees NIEM's evolution over time.

In addition to adding new NIEM types and properties to NIEM, it is possible to adapt existing external (non-NIEM) namespaces for use in the NIEM framework. This allows the use of external standards within NIEM IEPD, without requiring that the external standards themselves be NIEM-conformant (i.e. having the same naming and design rules). The intent here is to allow for the use of external standard components exactly as they were defined [39] .

NIEM is a well documented and supported standard. The NIEM program includes tools and support functions to help agencies at all levels of government take full advantage of this powerful data model [38]. Resources to support NIEM includes the following :

1. Webinars for training [40];
2. Tools Catalog to support NIEM IEPDs Development, Model Management / Search / Discovery [41];
3. Online training modules [42].

NIEM is not perfect. As noted in [43], the naming and design rules for NIEM are lengthy and difficult to comprehend, but simple cookbook implementations are possible without fully understanding those rules.

4.1.1 Maritime Enterprise Information Exchange Model

The Maritime domain of NIEM supports efforts for full Maritime Domain Awareness (MDA): the effective understanding of anything associated with the global maritime domain that could impact the security, safety, economy, or environment [44]. The National Information Exchange Model - Maritime (NIEM-M) XML vocabulary provides a combination of objects from NIEM core, the Maritime domain, and additions, via the Enterprise Information Exchange Model (EIEM) and IEPD, described later [45].

Through the definition of IEPD standards, the NIEM-M provides a common vocabulary for data and information exchange in five initial focus areas [46] :

- Position Exchange Model : Geospatial position, course, heading, speed, and status of a vessel at a given time. A series of position reports can be combined to produce track information.
- Indicators and Notifications Exchange Model : Indicators are used to inform or contribute to an analytical process. Notifications include warnings of a possible event and alerts about the execution of an event.
- Notice of Arrival Exchange Model : 96-hour advance notice that all vessels inbound to US ports are required to submit. This message names the vessel, lists the crew and passengers, and provides cargo information.
- Vessel Information Exchange Model : This package provides for the exchange of vessel information that is expected to remain constant over several voyages. This information includes things like the name and call sign of the ship.
- Consolidated Vessel Information and Security Reporting Exchange Model : It combines elements from the previous four IEPDs.

Figure 4.2 presents the Information exchange models built using these IEPDs.

The technical representation of XML is complex and it can be difficult to track the relationship of NIEM and NIEM-M components. To facilitate discussion and demonstrations with business and process managers, a graphical representation of each data model has been developed [47]. Those graphical logical models representation can be found on [46].

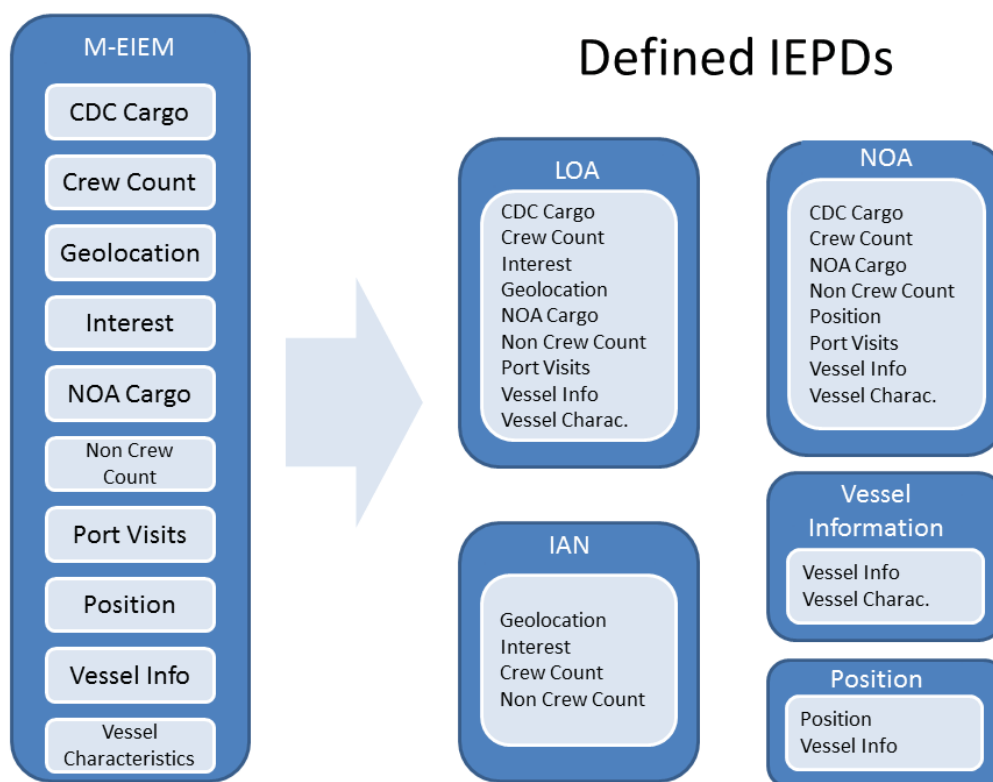


Figure 4.2: NIEM Information Exchange Models.

Leveraging the security markings, such as the attributes in NIEM-M based exchanges, a single data model can be used to support multiple versions of a product [47].

All these exchange packages enable someone to request the desired information using at least the attributes of a vessel and a geographical area of interest. The results can be constrained to a time window and filtered using the record metadata. These metadata include the security and classification level of the information to be returned. Each package also provides for the exchange of additional metadata specifically related to its particular objective.

In addition to the previously mentioned IEPDs, we can add the Department of Defence (DoD) NIEM Track Exchange Message, which is used by the DoD as an exchange model for a series of one or more geospatial positions captured over time that define the movement of a vessel [45] and which is built as shown in Listing 4.1. That listing provides the schema for a track exchange message in XML format.

Listing 4.1: Track Exchange Message

```

1 <xsd:schema targetNamespace="http://niem.gov/niem/domains/maritime/2.1/track/
  exchange/3.0" version="3.0">
2 <xsd:annotation>
3 <xsd:documentation>A vessel track</xsd:documentation>
4 <xsd:appinfo><i:ConformantIndicator>true</i:ConformantIndicator></xsd:appinfo>

```

```

5 </xsd:annotation>
6 <xsd:element name="Message" type="trkex:MessageType">
7 <xsd:annotation>
8 <xsd:documentation>A vessel track message</xsd:documentation>
9 </xsd:annotation>
10 </xsd:element>
11 <xsd:complexType name="MessageType">
12 <xsd:annotation>
13 <xsd:documentation>A data type for a single vessel track message.</
    xsd:documentation>
14 </xsd:annotation>
15 <xsd:complexContent>
16 <xsd:extension base="nc:DocumentType">
17 <xsd:sequence>
18 <xsd:element ref="mda:MessageIDURI" minOccurs="0" />
19 <xsd:element ref="mda:MessageModeCode" minOccurs="0" />
20 <xsd:element ref="trk:TrackSourceText" minOccurs="0" />
21 <xsd:element ref="mda:Vessel" minOccurs="0" />
22 <xsd:element ref="nc:ContactInformation" minOccurs="0" />
23 <xsd:element ref="mda:AdditionalRemarksText" minOccurs="0" />
24 <xsd:element ref="mda:Expansion" minOccurs="0" />
25 <xsd:element ref="mda:ICISMMarkings" minOccurs="0" />
26 </xsd:sequence>
27 </xsd:extension>
28 </xsd:complexContent>
29 </xsd:complexType>
30 </xsd:schema>

```

and where the positions are stored within the mda:Vessel element.

All the maritime IEPDs are built on the common business objects provided by the MDA EIEM. The EIEM are NIEM conformant (i.e. use the same naming and design rules), and provide common definitions for the core MDA objects, including [48]:

- Vessel Characteristics
- Vessel History
- Vessel Identification
- Movement
- Crew Nationality Count
- Non-Crew Nationality Count
- CDC Cargo
- Port Visit
- Interest
- Position
- Voyage Information

- Record Metadata

All IEPDs share these common object representations for interoperability. Each IEPD then defines only the additional necessary elements specific to that information product.

An Anomaly Report IEPD, which would contain all the metadata, should therefore be created in the same manner, reusing as much as possible. It should use a similar structure to that of the IEPDs that are already used within the maritime domain of NIEM. It is not necessary that all the fields be filled, either manually or automatically, when an anomaly report is created or shared. Naturally, metadata about the vessel on which the anomaly is flagged, the location of the anomaly and the timestamp at which it was detected must be in the data model. In addition, record metadata that relates to the releasability of the information (restrictions, classification level, etc.) must also be included. The next section will present the proposed NIEM-Maritime Anomaly IEPD, which can be used within SARA to describe an AIS-related anomaly.

4.2 NIEM Maritime Anomaly

Version 3.0 of NIEM-Maritime provides a metadata placeholder to describe an anomaly. Listing 4.2 presents the XML schema definition of this placeholder.

Listing 4.2: NIEM Maritime Anomaly Schema Definition

```
1 <xs:complexType name="AnomalyType">
2   <xs:annotation>
3     <xs:documentation>A data type for an out-of-the-ordinary occurrence.</
      xs:documentation>
4   </xs:annotation>
5   <xs:complexContent>
6     <xs:extension base="structures:ObjectType">
7       <xs:sequence>
8         <xs:element ref="m:AnomalyCategory" minOccurs="0" maxOccurs="unbounded" />
9         <xs:element ref="m:AnomalyContactInformation" minOccurs="0" maxOccurs="unbounded"
10          />
11         <xs:element ref="m:AnomalyDateTime" minOccurs="0" maxOccurs="unbounded" />
12         <xs:element ref="m:AnomalyDescriptionText" minOccurs="0" maxOccurs="unbounded" />
13         <xs:element ref="m:AnomalyLocation" minOccurs="0" maxOccurs="unbounded" />
14         <xs:element ref="m:AnomalyAugmentationPoint" minOccurs="0" maxOccurs="unbounded"
15          />
16       </xs:sequence>
17     </xs:extension>
18   </xs:complexContent>
19 </xs:complexType>
```

The elements of the complex AnomalyType object are defined as follows :

- Anomaly Category : A data concept for a type or kind of anomaly.
- Anomaly Contact Information : A point of contact (POC) for further information regarding an anomaly.

- Anomaly DateTime : A date and time an anomaly occurred.
- Anomaly Description : A description of an anomaly.
- Anomaly Location : A location where an anomaly occurred.
- Anomaly Augmentation Point : an Object Type that allows objects from multiple domains (Justice, Intelligence, etc.) to be attached to an object (in our case the Anomaly) without having to build a Maritime domain version of that object [49].

This structure is enough to accommodate the anomaly taxonomy previously developed. The taxonomy presented is composed of seven major categories (Error in AIS message, Conformity Issue, Duplication, Out of Date, Incompleteness, Inconsistency and Inaccuracy). These broad categories should be used in the Anomaly Category to enable a search / discovery of the information on them. Details of the anomaly should be placed in the description field and could be automated based on the structure of the taxonomy developed in chapter 3. As an example, an inaccuracy anomaly, such as the *GPS alteration*, can be described as *Inaccuracy - Inaccurate based on radar data - GPS alteration*, which follows the taxonomy. Additional remarks (from the operator for instance) could be placed within the Anomaly Augmentation Point. Listing 4.3 presents an example of a NIEM-M Anomaly .

Listing 4.3: NIEM Anomaly Subset Schema Definition

```

1 <m:Anomaly>
2   <m:AnomalyCategory>INACCURACY</m:AnomalyCategory>
3   <m:AnomalyContactInformation>...</m:AnomalyContactInformation>
4   <m:AnomalyDateTime>2016-01-20T10:27:03.522Z</m:AnomalyDateTime>
5   <m:AnomalyDescriptionText>Inaccurate based on radar data - GPS alteration
6   </m:AnomalyDescriptionText>
7   <m:AnomalyLocation>
8     <m:LocationAugmentation>
9       <m:LocationPoint>
10        <gml:Point gml:id="anomaly_point1">
11          <gml:pos>-1.1 -1.1</gml:pos>
12        </gml:Point>
13      </m:LocationPoint>
14    </m:LocationAugmentation>
15  </m:AnomalyLocation>
16  <m:AnomalyAugmentationPoint>Additional elements can be included here for
    remarks from the operator</m:AnomalyAugmentationPoint>
17</m:Anomaly>

```

As shown, the currently defined NIEM-Maritime anomaly structure can be used to describe all the anomalies in the taxonomy. The next section will present the design of a complete anomaly exchange report as well as the anomaly request logical model.

4.3 Record and Security Metadata

Record and security metadata are needed because constraints on sharing information exist. There will always be some restrictions on sharing sensitive operational or personal information. In addi-

tion, foreign partners, governments, and the private sector may impose limits on use or dissemination of their information products [50]. Metadata within SARA must respect these realities and provide a responsible means to share information. Our proposed NIEM-based anomaly exchange model must tag the data, identifying and authenticating users.

As mentioned in [50], most information authorization models are limited to access controls defined and enforced at the network or application-level, rather than at the data-level using inherent characteristics of specific information resources. As shared services are adopted, however, access controls will have to be applied to the data itself using tags. Information tagging is an approach where standard attributes (tags) are attached to a piece of information to describe it. While manual discovery and access capabilities benefit from information tagging by guiding users directly to specific information based on their profile, it also can enable automated enforcement of access decisions. By matching the user attributes with corresponding information attributes, the automated delivery of information is improved along with the security and protection of that information from inappropriate recipients. Information tagging further assists in meeting records management requirements, responding to disclosure inquiries, integrating privacy protections, and remediating erroneous data disclosures and modifications.

The maritime domain of the NIEM and its extension provide a means to perform information tagging using the entitlement markings for the IEPD instances (an XML file that contains data is called an *instance document*). Generally speaking, entitlement marking is the association of a set of attributes to a specific piece of information that defines the information protection, source and other metadata about that piece of information [51].

The anomaly reports should adopt the same entitlement method, which is currently used in the Maritime Information Sharing Environment (MISE) in the United States and has a proven track record. The block metadata also contains the Intelligence Community Metadata Standard for Information Security Marking (IC-ISM), which is one of the Intelligence Community (IC) Metadata Standards for Information Assurance and is the preferred way to apply information security markings within XML document instances.

In addition to the access restriction, it is envisaged that the information provided by SARA may be incomplete, vague, or inaccurate. Metadata that help stakeholders to assess the provenance of information, such as contact information, is essential for ensuring quality control.

The reader should refer to [51] for a complete description and example of entitlement marking within an IEPD document instance.

4.4 NIEM-based Anomaly Exchange Report

Note that because of the high level of reuse in NIEM, some elements that would be allowed by the XML schema definition might never be used in practice. For example, the NIEM XML schema definition allows a birth date to be specified for any person, but this would only be specified if the person is a crew member, not a document creator. Implementers of the schema should be schema conformant, but expect that only a subset of the elements made available by the schema will actually be exchanged. Listing 4.4 presents the schema definition of the proposed IEPD.

Listing 4.4: NIEM Maritime proposed Anomaly IEPD

```

1 <xsd:element name="Message" type="anoex:MessageType">
2   <xsd:annotation>
3     <xsd:documentation>A vessel track message</xsd:documentation>
4   </xsd:annotation>
5 </xsd:element>
6 <xsd:complexType name="MessageType">
7   <xsd:annotation>
8     <xsd:documentation>A data type for a single vessel track message.</
9     xsd:documentation>
10  </xsd:annotation>
11  <xsd:complexContent>
12    <xsd:extension base="mda:DocumentType">
13      <xsd:sequence>
14        <xsd:element ref="mda:Vessel" minOccurs="0" />
15        <xsd:element ref="m:Anomaly" minOccurs="0" />
16      </xsd:sequence>
17    </xsd:extension>
18  </xsd:complexContent>
19 </xsd:complexType>

```

which, when expanded, is represented by figure 4.3.

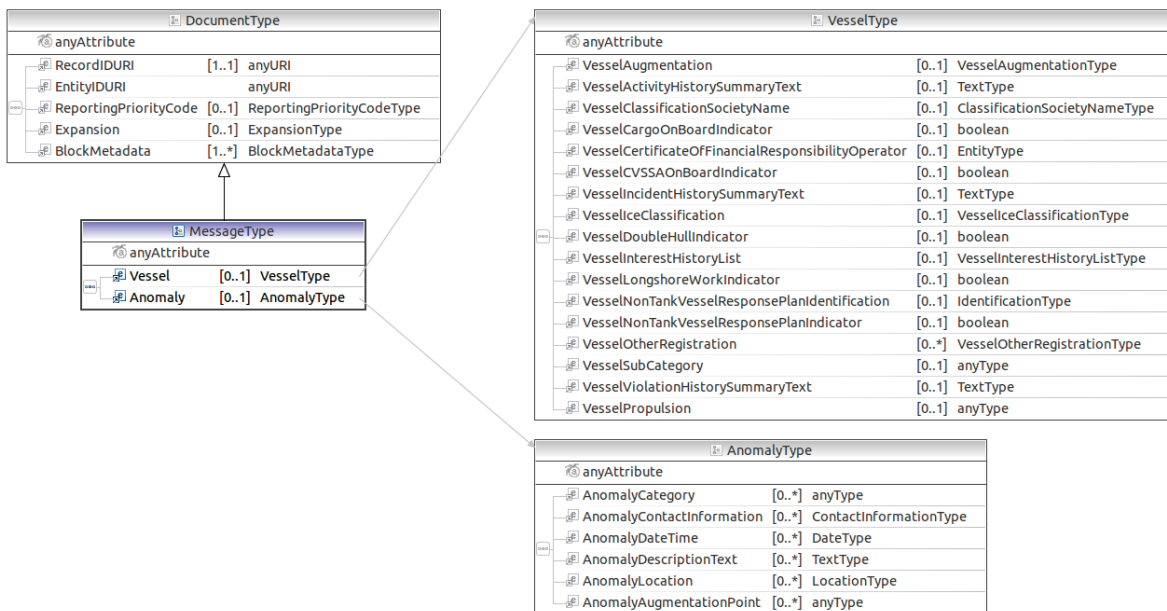


Figure 4.3: Anomaly Exchange XML Schema Definition (XSD).

It is important to note that an image of the vessel can be added to this anomaly report under the VesselAugmentation element, which already contains a VesselImage placeholder. If additional images or video are to be added, those should be included under the AnomalyAugmentationPoint, preferably in the form of a URL that can be accessed when the anomaly report reaches the requester of the information.

4.5 Search Parameters for an Anomaly Report

Several requests are already defined as part of the maritime domain of the NIEM (see [52] for an example). The anomaly request logical model should be based on these already adopted schemes. As depicted in the logical diagram in figure 4.4 one should be able to search for anomaly reports within a geospatial area and time window, constraining the results to specific vessels or anomaly types.

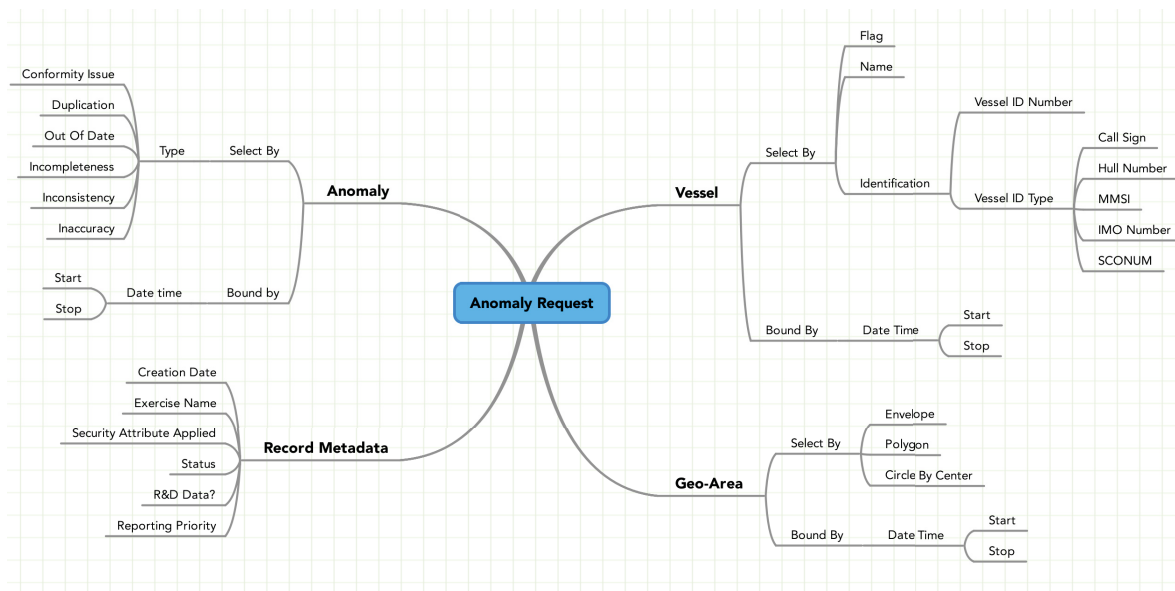


Figure 4.4: Anomaly report request parameters.

Results of a request will be filtered out based on the security metadata and the level of security of the requester (see figures 4.5 and 4.6, acronyms are defined in Table 4.3).

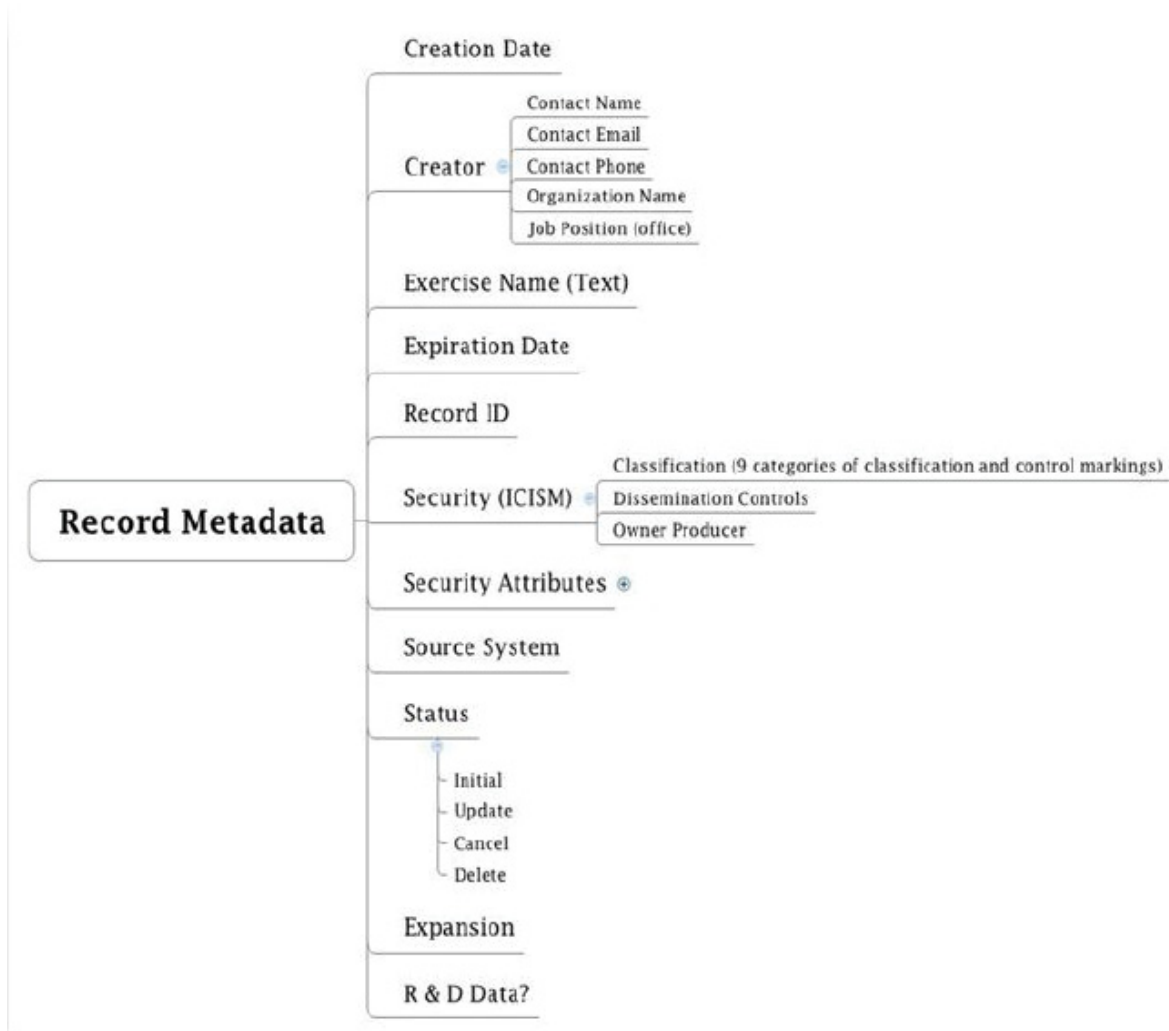


Figure 4.5: Record metadata element.

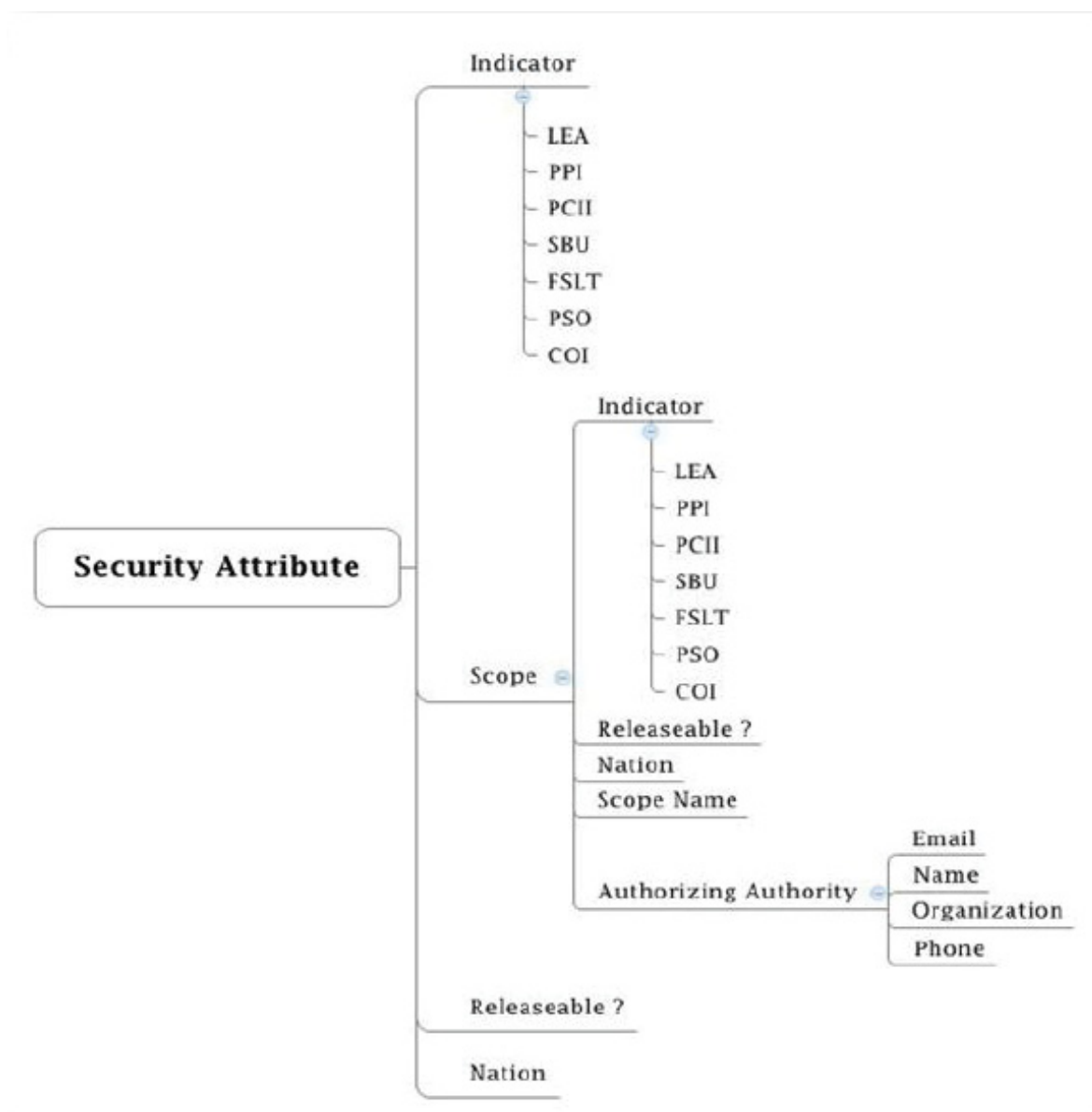


Figure 4.6: Security attributes element.

Record metadata is included in a report message to provide additional information for handling and/or managing the message. In figure 4.4, the record metadata depicted on the left side of the graphic includes such information as record status, creation date, and exercise name. In some cases it may be useful to further filter the result set on the client side by record metadata. For example, filter for only updated records, only records created in the last 24 hours, or only records associated with a specific exercise.

4.6 Data Element Description

This section describes the elements part of the proposed Anomaly IEPD metadata.

4.6.1 Vessel Identification

Table 4.1 presents the description of the vessel identification data elements.

Common Name	Description
Flag	The national flag under which a vessel sails
Name	The name of a vessel
Registered Owner	<p>The entity that owns a vessel. Information includes:</p> <ul style="list-style-type: none"> • Name : The name of the person that owns a vessel • Nation : The name of the nation that owns the vessel
Identification	<p>Unique identification attributes of a vessel are:</p> <ul style="list-style-type: none"> • Vessel ID Number : The number of a vessel • Vessel ID Type : Vessels can be identified by the following information: <ul style="list-style-type: none"> – Call Sign : The call sign for a vessel – Hull Number : The hull number of a vessel – MMSI : MMSI of a vessel – IMO : Number The International Maritime Organization Number (IMO number) of a vessel – SCONUM : The Ship Control Number (SCONUM) of a vessel – Official CG Number : An official United States Coast Guard Number (USCG Official Number) of a vessel – Other Registration : Other registration information of the vessel <ul style="list-style-type: none"> * Type : Information about another, not otherwise specified registration for a vessel * Number : Registration number of a vessel

Table 4.1: Description of the vessel identification data elements.

4.6.2 Anomaly Identification

Table 4.2 presents the description of the anomaly identification data elements.

Common Name	Description
Anomaly Category	A data concept for a type or kind of anomaly.
Anomaly Contact Information	<p>A point of contact (POC) for further information regarding an anomaly. Information includes:</p> <ul style="list-style-type: none">• Contact Name : The name of a person who is the contact for the anomaly creator• Contact Email : An electronic mailing address by which the anomaly creator may be contacted• Contact Phone : A telephone number by which the document anomaly may be contacted• Organization Name : The name of an organization who is the anomaly creator• Job Position (Office) : Brief description of a position
Anomaly Date-Time	A date and time an anomaly occurred.
Anomaly Description	A description of an anomaly.
Anomaly Location	Geography Markup Language (GML) point representation for the position.
Anomaly Augmentation Point	Additional information.

Table 4.2: Description of the anomaly identification data elements.

4.6.3 Record Metadata

Table 4.3 presents the description of the record metadata elements.

Common Name	Description
Creation Date	A date a document was created
Creator	<p>Entity primarily responsible for creating the content of the resource. Information includes:</p> <ul style="list-style-type: none"> • Contact Name : The name of a person who is the contact for the anomaly creator • Contact Email : An electronic mailing address by which the anomaly creator may be contacted • Contact Phone : A telephone number by which the document anomaly may be contacted • Organization Name : The name of an organization who is the anomaly creator • Job Position (Office) : Brief description of a position
Exercise Name	Exercise Name
Expiration Date	A date a transmitted document expires
Record ID	A unique ID identifying a record
IC-ISM Marking (Security)	<p>Security classification of data. Information includes:</p> <ul style="list-style-type: none"> • Classification : A single indicator of the highest level of classification applicable to an information resource or portion within the domain of classified national security information • Dissemination Controls : One or more indicators identifying the expansion or limitation on the distribution of information • Owner Producer : One or more indicators identifying the national government or international organization that have purview over the classification marking of an information resource or portion therein

Common Name	Description
Security Attributes Applied	<p>Characteristics of a persona that defines a user in a particular role. Security Attributes include:</p> <ul style="list-style-type: none"> • Indicator : Indicates the security that must be applied to the record <ul style="list-style-type: none"> – Law Enforcement (LEI) – Privacy Protected (PPI) – Sensitive But Unclassified (SBU) – Protected Critical Infrastructure Information (PCII) – Federal/State/Local/Tribal/Territorial (FSLT) – Private Sector Only (PSO) – Community of Interest (COI) • Scope : Indicates the security that must be applied to this record within the context of the named scope <ul style="list-style-type: none"> – Indicator : LEI, PPI, SBU, PCII, FSLT, PSO, COI – Releasable : Whether the data is releasable in the context of the names scope – Nation : The nations which the data can be released in the context of the named scope – Scope Name : Name of the scope in which the security and releasability is modified – Authorizing Authority : Entity authorizing the use of the set of attached scope attributes on this metadata element. Includes the following information: <ul style="list-style-type: none"> * Email : The email of the authorizing authority * Name : The name of the authorizing authority * Organization : The organization of the authorizing authority * Phone : The phone number of the authorizing authority • Releasable : Whether the data is releasable • Nation : Nations allowed for release

Common Name	Description
Status	<p>A status of a record that can be any of the following:</p> <ul style="list-style-type: none"> • Initial : New record • Update : Change or update to an existing record • Cancel : Cancellation of existing record • Delete : Deletion of an existing record
Expansion	Additional information.
Research and Development Data	True if the information in the record is for research and development efforts, false otherwise
Reporting Priority	Numeric value representing how often this record is updated

Table 4.3: Description of the record metadata elements.

4.7 Suspicious Activity Reporting and its Lessons Learned

Using existing IEPDs from other domain provides possible leads for the representation and the design of metadata related to an anomaly. This section briefly describes the Suspicious Activity Report (SAR) IEPDs and the lessons learned during their development to ease the adoption of a standard reporting format across multiple agencies.

A SAR is sent by an agency identifying suspicious activity to a Fusion Center and various other interested agencies. Figure 4.7 represents the components of the IEPD associated with the SAR.

Note that all the information attached to a person is defined by default in the core components of the NIEM. All documentation related to this IEPD can be found here [53]. What is most interesting is the Findings and Recommendations of the SAR Support and Implementation Project [54]. From it, one can extract the following lessons learned that broadly apply to an acceptance of SARA within the community as a tool for anomaly reporting and sharing. The lesson learned are summarized here :

- Strong executive leadership is an essential element leading to the success of a program (in this case the SAR). Agencies/Department should educate and gain the support of policymakers.
- It is important to share information, as opposed to stockpiling it. Accepting the mind-set that information must be shared in order to make this program successful is an essential philosophy that all policymakers must adopt.

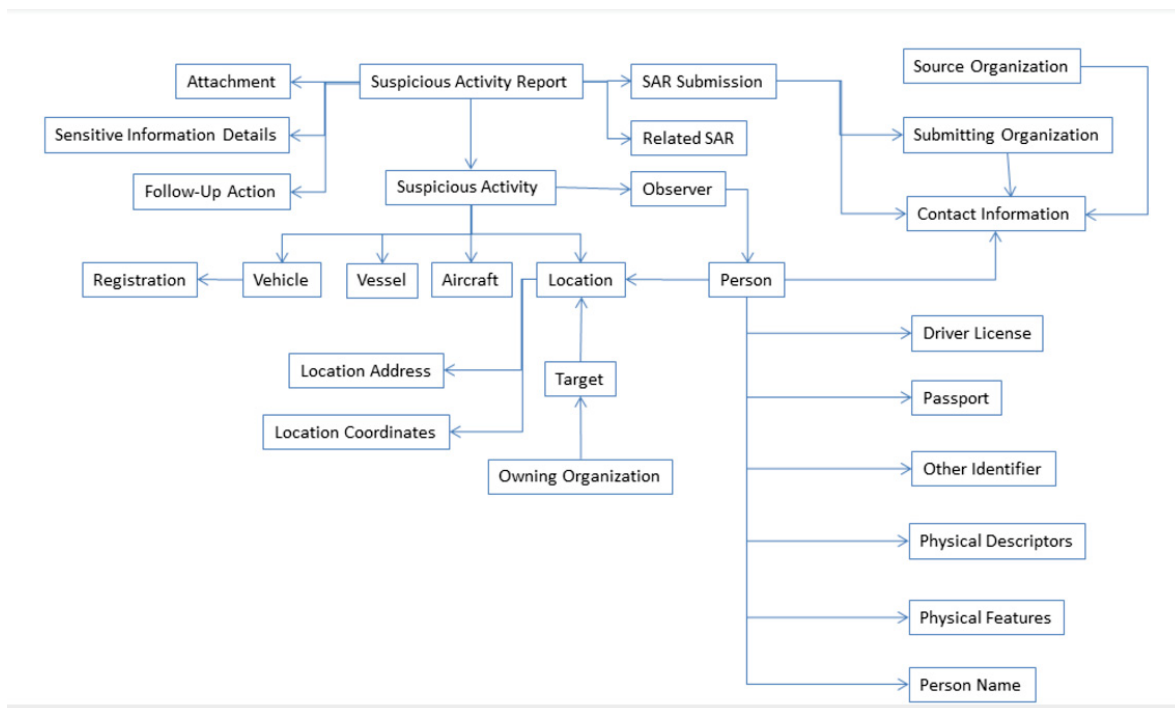


Figure 4.7: Suspicious Activity Report design.

- The gathering, processing, reporting, analyzing, and sharing of information is critical.
- Training is a key component of the process, as such, all relevant agency personnel must be trained to recognize behavior of interest.
- There is a need for a common national methodology for the sharing information.
- A defined process is needed by the originating agency to ensure that anomaly reporting is made available to fusion centers in a timely manner. Incorporating outreach to the public, law enforcement, and the private sector in the collection process is important to the success of the program.
- In order to leverage resources and avoid duplicated efforts, agencies should use existing information technology, common systems, and information sharing relationships, so that information can be shared more broadly and effectively.
- Technology and use of common national standards enhance the capability to quickly and accurately analyze data in support of controlling and preventing criminal activity (in the case of the SAR).

Part 5

Requirements

This section introduces a list of requirements on which the design of the overall system should be based. These requirements are based on the use cases presented in section A. Of course, this list is not definitive and modification to it will most probably occur, if such a system should be implemented. It should be considered as a basis for discussion and will evolve during development as implementation progresses.

This section is organised as follows :

- Section 5.1 presents the external interface requirements.
- Section 5.2 presents the software interfaces.
- Section 5.3 presents the system feature requirements.
- Section 5.5 presents other system requirements.
- Section 5.5 presents performance and security requirements.

5.1 External Interface Requirements

It is to be understood that for any integration within existing C2 systems, the design and implementation of the proper user interface components should be left to the owner of the said system. In this case, for anomaly submission or retrieval, SARA provides a well documented RESTful Application Programming Interface (API) (see next section 5.2 for details on REST). Systems owners would be responsible for implementing the user interface for searching, submitting and visualizing the anomalies.

In any implementation of SARA, some tasks would have to be carried out by an administrator, or by the ruling authority (see section 5.3.4 for a description of the types of users), and thus would require basic web interfaces. User profiles creation and editing as well as retrieving and editing ship profiles, exemplify this requirement. To this end, very simple interfaces that send the correct query to SARA have to be made available to an administrator.

5.2 Software Interfaces

In an effort to make SARA as available and scalable as possible, it is to be implemented as a RESTful web service (from Representational State Transfer (REST)). RESTful frameworks have become standard for web services, and, as such, are advantageous given that SARA is to coordinate with existing and future services and applications.

A RESTful service implies a client-server architecture, where the server interfaces between the users and the database. Although the user population size is not a primary concern, the server design should take into account the potential need to satisfy a large community of users. Therefore a non-blocking, scalable framework, such as Node.js, is strongly recommended. The server is also responsible for user creation and validation.

To maximise its chance of interoperability with other systems, the anomaly exchange reports should be formatted in XML under the NIEM framework. The use of this standard is likely to ease its adoption within the community, as its support is currently growing.

The functional requirements are presented in the list below:

- REQ-1. SARA shall be highly available and scalable.
- REQ-2. SARA shall implement a RESTful API, using a non-blocking, scalable framework.
- REQ-3. SARA shall return the requested anomalies in XML format, conforming to the NIEM schema.

5.3 System Features

This section describes the main system features requirements.

5.3.1 User Reporting of Anomalies

Security agents can identify and submit ship anomalies through SARA. Anomaly reporting needs to be independent of a User Interface (UI) because SARA should not be tied to any one single system but function as an independent service. The service defines the common toolset used by all third-party applications communicating with SARA.

Once a security agent identifies an anomaly, he enters the required information in a report and sends it to SARA for validation and storing. It is strongly advised that data from existing AIS reports should be used to pre-fill anomaly reports. If SARA is integrated into an existing system, it can leverage the available data and compare it to the available AIS reports, making reliable suggestions on how to fill the anomaly report.

Barring any attestable data, the anomaly characterization is left entirely to the user, although some validation will be enforced within the submission form fields. To that end, a well documented web service endpoint description is mandatory.

Also, to enforce a common language in the reporting of anomalies, SARA should enable one to retrieve the taxonomy of anomalies. Users should use this taxonomy when implementing their interfaces for anomaly reporting.

The functional requirements are presented in the list below:

- REQ-4. Anomaly reporting shall be a service independent of a UI.
- REQ-5. The user shall specify a ship for anomaly reporting. He shall only report one ship at a time.
- REQ-6. SARA shall pre-fills the anomaly report with as much data as possible.
- REQ-7. The user shall manually confirm input fields that are relevant to the specified anomaly type. For example, if the reported destination is obviously incorrect given an observed ship's trajectory, then the user would be required to confirm that the reported destination is wrong and might even enter a new destination, if one can be inferred.
- REQ-8. SARA shall perform error checking on input field in order to minimize human error. For example, the heading cannot be greater than 360 degrees.
- REQ-9. SARA shall uniquely identify generated anomalies, by combining the erroneous field, timestamp, and location, so as to facilitate the identification and reduction of duplicate anomaly reports from multiple sources.
- REQ-10. SARA shall provide a web service endpoint to retrieve the anomaly taxonomy.

5.3.2 Retrieval of Anomalies

SARA's users can query the web service for anomalies by specifying certain attributes, such as a time window and bounding box. The queries available are defined on the SARA server and made available in a RESTful API accessible through Hypertext Transfer Protocol (HTTP). If integrated within a UI, some queries can be performed autonomously. A border patrol ship could, for example, query information on nearby ships as soon as they come into range of sensors. Queries into the anomaly history of nearby ships would be submitted without user input by the on-board ECDIS software.

The functional requirements are presented in the list below:

- REQ-11. SARA shall be able to return all anomalies inside a physical bounding box.
- REQ-12. SARA shall be able to return all anomalies of a specified anomaly type (Eg. Conformity, Duplication, time window, etc.).
- REQ-13. SARA shall be able to perform queries on any combination of filters applied on its metadata.

5.3.3 Retrieval and Edition of Ship Profiles

Anomalies should also be attached to particular vessels. A vessel with all its related information and associated anomalies is referred to here as ship profile. It is anticipated that a single vessel might be represented by multiple profiles, due to errors in AIS data (misspelling, errors in MMSI and so on), therefore creating an ambiguity on some vessels. As such, SARA must provide a way for a user to retrieve those profiles and edit them, in order to remove potential ambiguity from the system and provide accurate information to its users.

The functional requirements are presented in the list below:

- REQ-14. SARA shall create a new ship profile for an anomaly, when such anomaly cannot be attached to an existing one.
- REQ-15. SARA shall allow administrators to retrieve and edit ship profiles through a simple user interface.
- REQ-16. SARA shall allow administrators to resolve ship profiles ambiguities through its RESTful API.

5.3.4 User profiles

SARA incorporates different user types with varying levels of access to its components. The following example user types are ordered by rank from highest to lowest.

- **Administrator** : The administrator benefits from the highest level of access to the system with the purpose of performing maintenance jobs on the database. This user also fills the role of an identity provider, and thus creates and edits user profiles.
- **Ruling Authority** : This user can edit all anomaly reports and has access control over other user accounts of lower tier. This user maintains the quality of reports and can discard reports that are deemed abusive or irrelevant. The ruling authority can also intervene to issue warnings or ban users who infringe on operational rules.
- **Registered User** : A registered user can generate and view anomaly reports after creating a profile. These include submitting users and maritime security professionals. A submitting user may record an observed anomaly for a given ship, in order to preserve and amass knowledge about it. Maritime security professionals, a category that includes Navy officers, Coast Guard operators and MSOC navy operators, could use SARA to consult and view AIS-related anomalies associated with their maritime interests. They would increase their work efficiency through SARA by focusing their attention on the ships most deserving of attention.

To simplify user creation and user profile edition, SARA should provide a simple user interface to an administrator.

In addition, the information contained within SARA might have different level of classification. For instance, PAL is producing Protected B information products. Therefore, SARA must restrict access to its information. Also, under the NIEM framework, SARA will have to support NIEM-based information entitlement (see section 4.3 for more details).

The functional requirements are presented in the list below:

- REQ-17. SARA shall handle user profiles with different privileges.
- REQ-18. SARA shall provide access restriction to its products with regards to their classification levels.
- REQ-19. SARA shall allow administrators to create and edit user profiles through a simple user interface.
- REQ-20. SARA shall allow users to set their password, as long as it respects a few security constraints.

5.4 Other System Features

This section lists other system features requirements.

5.4.1 Autonomous Generation of Anomaly Reports

SARA requires a processing unit that is designated to the detection of anomalies in a large database. It is envisioned that SARA would be implemented with an automatic anomaly detection component. This component should connect to an existing AIS repository, such as MSARI, in order to process anomaly reports from its vast amounts of recorded data. It is also possible that SARA might use a third-party system, TimeCaster for instance, to get AIS-related anomalies.

The functional requirements are presented in the list below:

- REQ-21. SARA shall allow connections to external databases for processing of anomalies.
- REQ-22. The automatic anomaly detection component shall be a fire-and-forget process with optional boundary settings.
- REQ-23. The automatic anomaly detection component shall accept time window and geographic boundaries.
- REQ-24. This process shall not negatively impact the responsiveness of user interaction with SARA.
- REQ-25. Metadata concerning reporting the source and organisation shall be sent for traceability.
- REQ-26. SARA shall allow connections to a third party system reporting AIS-related anomalies.

5.4.2 Performance Tracker

Because SARA is expected to be made up of many components with varying processing and communication loads, each of which could negatively impact the others, it is essential to track their status and efficiency. This system will allow administrators to make more surgical interventions when something goes wrong.

The functional requirements are presented in the list below :

- REQ-27. SARA shall include a performance tracker to monitor the server connections between SARA's components as well as user connections.
- REQ-28. SARA's performance tracker shall also monitor the processing workload.
- REQ-29. SARA shall warn the user when connection issues appear, so that the user may fix the problem if it is at his end or at least curb his expectations of responsiveness.

5.4.3 Pre-processing watchdog

In high-level terms, this watchdog exists to guarantee the responsiveness of all queries available to SARA. Some query products require a lot of processing power and time to produce. Since millions of AIS reports are generated each day, it makes sense to consider how data will be pre-processed in order to guarantee the responsiveness of user queries.

This component is loosely defined, as it will likely evolve significantly during development. It is therefore helpful to consider an example. Say a reliability rating is to be implemented for all ships. This reliability might be based on weights assigned to all reported anomalies relative to their severity and frequency of occurrence for each given ship. When an operator wishes to view all the information available for ships in his vicinity ordered by reliability ratings, that reliability rating must be pre-computed. The computation must be handled in batches, as these calculations can be very Central Processing Unit (CPU) intensive. It is also worth noting that the result of these products may not be up to date, due to the nature of real-time data streams, as new data keeps arriving while the existing data is worked on. If the desired value is critical, a combination of pre-computed data and live calculation is necessary to get the most up-to-date value.

The process of pre-computation should start after a set delay, if any new data is available or after a quota of incoming data reports has been reached.

The functional requirements are presented in the list below:

- REQ-30. SARA shall implement a watchdog that tracks incoming data and computes pre-defined products.

5.5 Other Nonfunctional Requirements

The list below presents other nonfunctional requirements:

- REQ-31. User interactions with SARA must be as responsive as possible. Evidently, this requirement is loosely defined and performance can vary greatly due to the available bandwidth in a real environment. As the most common SARA user is expected to be a security officer with many other duties already assigned to his role, it is important that SARA doesn't artificially add downtime to an already busy schedule.
- REQ-32. The processing and acquisition of new anomalies must keep up with the influx of new AIS reports at all times. In this sense, SARA must perform as a quasi-real-time system. The danger is that too much data comes into SARA and the acquisition becomes a bottleneck for the system, eventually overflowing the input buffers and crashing the application.

5.5.1 Design and Implementation Constraints

In order to guarantee high availability and redundancy of SARA, multiple architecture and hardware constraints must be considered. These requirements, however, can only be defined when lower level implementation decisions, such as SQL versus NoSQL (see section 6.3.7) have been made. A MongoDB production cluster (NoSQL) is used here as an example. Its requirements are as follows:

- **Two or more Replica Sets as Shards** : A replica set in MongoDB is a group of mongod processes that maintain the same data set. Replication serves both as data redundancy and as a means of increasing read capacity from clients. Each member of the replica set should be on a separate machine.
- **One or more query routers** : The query routers interface with the application driver and redirect queries from clients. If heavy client traffic is expected, the number of routers can be increased to balance the load.
- **Three configuration servers** : Configuration servers store metadata for a shared cluster. A production shared cluster requires exactly three config servers exclusively assigned to the task.

The minimal production environment therefore requires 6 separate machines. See [55] for a complete description of requirements and deployment guide for a MongoDB production cluster.

This page is intentionally left blank.

Part 6

Design

This section presents the SARA's software design concepts and is organized as follows:

- Section 6.1 presents the high-level architecture.
- Section 6.2 details the design of the components of the application server.
- Section 6.3 presents the design of the anomaly database and data structure.
- Section 6.4 presents the functionality of processing modules.
- Section 6.5 provides the design rationale.
- Section 6.6 shows an example of integration of SARA within an information sharing environment based on NIEM.

6.1 Architectural Design

From a high-level point of view, users can perform actions that fall into two major categories:

- Send anomaly reports to SARA.
- Retrieve anomaly reports from SARA.

SARA presents a similar structure to a usual web service provider. Looking at the Figure 6.1 below:

1. Users interact with a client system trusted by SARA, such as a web application, local application, or an existing ECDIS. SARA clients are both consumers and providers of data; therefore, all communication between components is bidirectional. More detail on the client side of SARA is out of scope at this point in the project. Focus is instead aimed at making

SARA's communication with clients as adaptable as possible in order to maximize interoperability with existing NIEM frameworks. Among other things, this implies the adoption a RESTful messaging structure, which has become the de facto standard in lightweight message transfer architecture, and would enable SARA to readily interconnect with existing and future NIEM networks.

2. The application server handles user requests and responses to and from the database. In this, the application acts primarily as a mapper between service requests (in XML or JavaScript Object Notation (JSON) code) to database language. Conceptually, SARA makes no distinction between individual users and trusted systems. Security restrictions are specified for each client system; however, each system in turn may disseminate consumed data to many users.
3. The database stores anomalies, ship profiles, user profiles as well as other metadata for database maintenance.
4. Finally, the processing component handles all validation and computationally expensive calculations of anomaly reports and sends the resulting products back to the database. These calculations may be performed in a batch process and include, among other things, the reliability rating of all ships or the creation of ship profiles from a trusted database. A secondary objective of the processing component may be to interface with AIS streams or databases and generate anomaly reports autonomously.

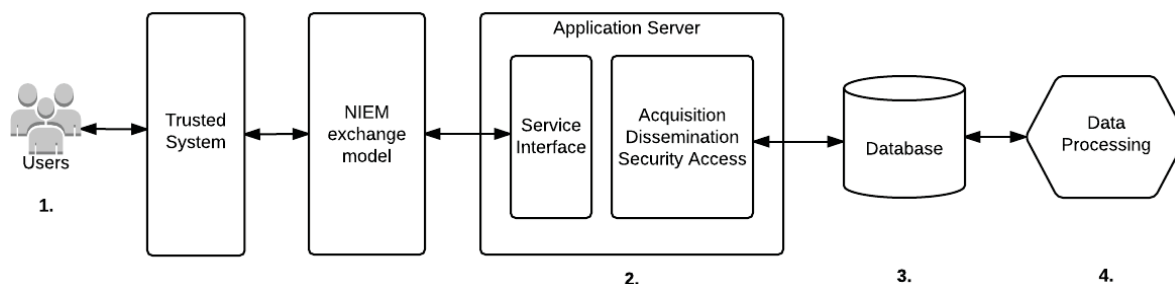


Figure 6.1: A high level overview of SARA's architecture. 1. Users are expected to interact with SARA through a trusted system such as a web application or other ECDIS. This client system provides the human interface (the UI) required to help the user build, send and receive queries to and from SARA. 2. The application server processes requests from users through the service interface and maps them to the database. 3. The database stores anomalies, ship profiles, user profiles and metadata. 4. The data processing unit is in charge of calculations and algorithms required to support the database. These include the calculation of reliability indexes for ships, creation of ship profiles and other routine maintenance processes.

6.2 SARA Application Server

The application server's primary purpose is to convert service commands to database queries. The set of available commands is pre-defined and discrete. The list of available queries must

therefore be made available to the user. The objective is to cover the most common cases used by operators while ensuring that the process for adding new queries is streamlined. When the user sends a request to the server, the query name must be recognized and its required attributes (when available) must be specified for the query to be mapped to the database. The methods for handling these queries, as well as all the processes required to support them, form the domain logic of the system. Housing all the domain logic on the server is desirable in order to decouple it from other tasks, decrease maintenance and facilitate testing. However, this separation is not necessarily realistic, depending on the database structure and efficiency requirements. Indeed for relational databases, shifting some of the logic to the database in the form of stored procedures often increases productivity.

Figure 6.2, shows the major functional components housed by the application server in two major groups. The service layer handles the communication with clients while the domain layer handles all the logic associated to core functionalities of the system (anomaly reporting and submission, ship profile reporting and editing, etc.). The following subsections provide more detail on the subcomponents within the application server, as shown in Figure 6.2.

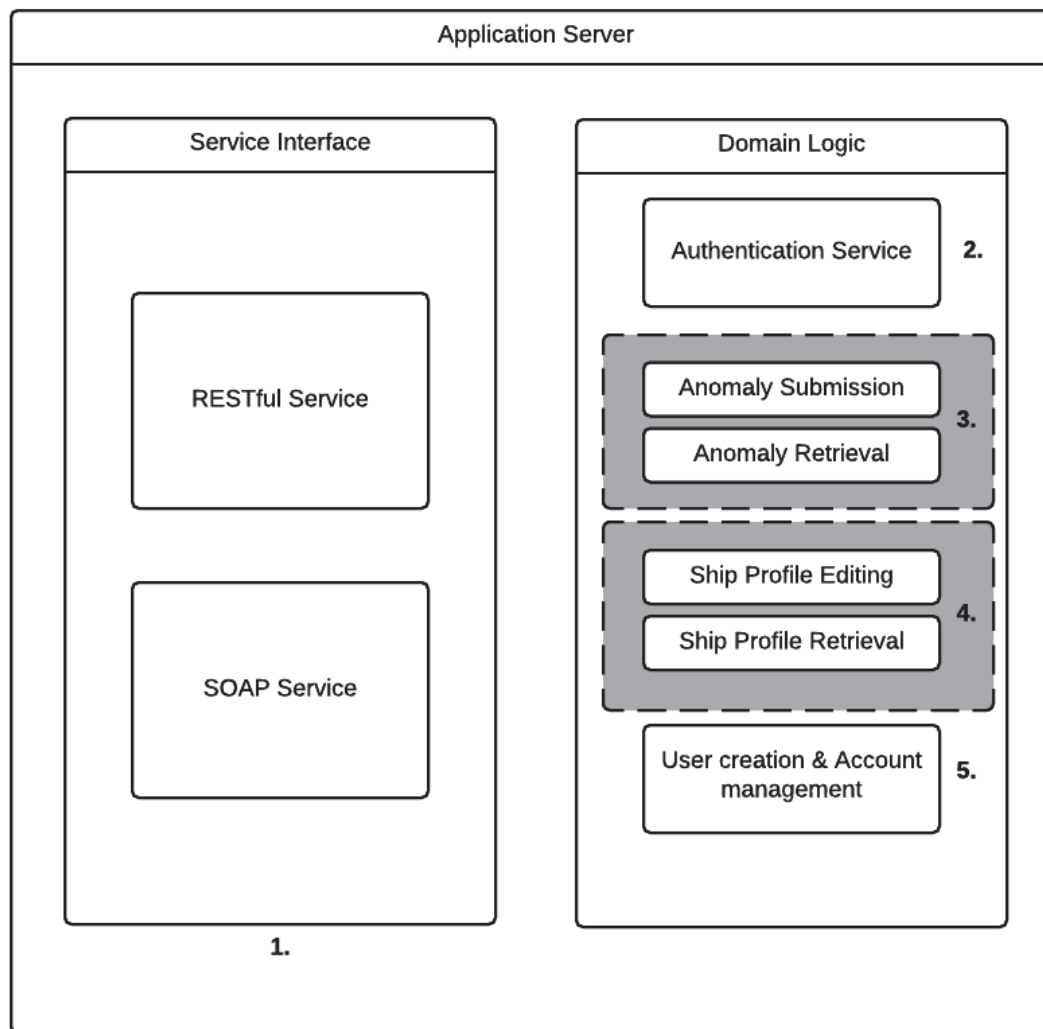


Figure 6.2: Inside the application server. 1. All communication with clients goes through the service interface either through REST or SOAP messages. 2. The authentication service allocates each user a timed token after a successful login. Every subsequent query needs to also send the token in order for the request to be valid. 3. The core purpose of the application server is to map incoming anomaly requests into database queries. These queries include both submissions of new anomalies as well as requests for stored anomalies. 4. Operators may also edit ship profiles through queries or request to look at stored ship profiles. 5. The creation and management of user accounts also occurs through the service interface but requires different access rights. Only designated administrators (See Section 5.3.4) can create new accounts and edit security level access for example. Other user profile details can be edited by the account owner.

6.2.1 Anomaly Submission

The anomaly submission and retrieval (see item 3, Figure 6.2) form the basis of SARA's capabilities allowing operators to record their work for posterity and make use of their colleagues' records to

increase their own productivity. To start off the anomaly submission, an operator fills in the details of an observed anomaly through a trusted system which then sends a report to SARA. The list below (accompanied by Figure 6.3) details the order of operations at the receiving end.

1. SARA first performs a series of validation operations, ensuring the input fields are correct and returning error specific reports otherwise.
2. In the case where the anomaly report doesn't contain any information on the suspect vessel, the anomaly is sent directly to the database. The anomaly report's value is diminished by not directly referencing a vessel but may still be useful in the long term and is therefore persisted.
3. When the suspect vessel information is available, the reported ship is identified by looking at the information contained in the anomaly report and searching for a matching vessel in SARA's ship database. Note that identifying a ship from the anomaly report data is a potential risk both in terms of likelihood of success as well as in terms of performance (section 7.1 covers the issue in greater detail).
4. If no matching ship is found in the database, a new ship profile is created from the available report data. Relying on reported data for the creation of new vessel profiles is a risk that can lead to duplicate profiles (Section 7.2 covers the issue in greater detail).
5. The anomaly is recorded on the ship profile.
6. The anomaly report is stored in the database.
7. A successful transaction message is returned to the operator.

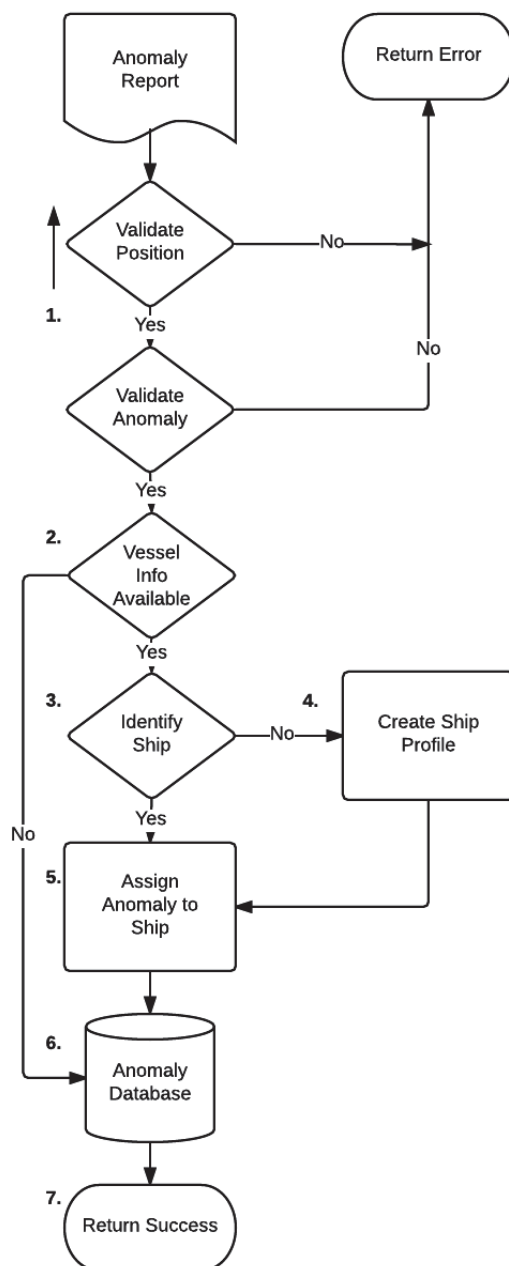


Figure 6.3: The best case scenario flowchart for submission of new anomaly reports. 1. Validation with error specific messages. 2. Immediate storing of anomalies not containing information identifying the suspect vessel. 3. The identification of a ship in SARA’s database from anomaly report information. Note that this is a complex process and risky in terms of likelihood of success, since there may be many matching ship candidates with no clear singular result, and in terms of performance since the search procedure may be very computationally expensive. 4. The creation of a new ship profile when a matching ship profile was not found in the database. As this process is automated, there is a risk it will create duplicate ship profiles from reports containing spelling errors or other input discrepancies. 5. This process links an anomaly report to a ship profile. 6. The anomaly is stored in the database. 7. A successful transaction message is returned to the operator.

6.2.2 Anomaly Retrieval

Anomaly retrieval requests (see item 3, Figure 6.2) allow the operators to obtain a subset of anomalies based on specified filters. Anomaly retrieval starts off with a request by the operator defining the filters for extraction. The list below and the accompanying Figure 6.4 show the operation flow on handling the request:

1. Input validation ensures the request fields respect requirements.
2. Before the query is executed, the operator's security level is fetched from his/her profile. Security specifications are an integral part of NIEM and require the query process to take into account the operator's security level when searching through the database. Through NIEM record metadata, only database entries that match the operator's security level are traversed. The alternative scheme would be to execute the query regardless of security levels and filter the result set before returning it to the operator. However, that would imply a lot of potentially wasted work when a query is computationally expensive and the operator's security level isn't eligible to see most of the results.
3. The result set is formatted for output, in JSON or XML and sent to the user.

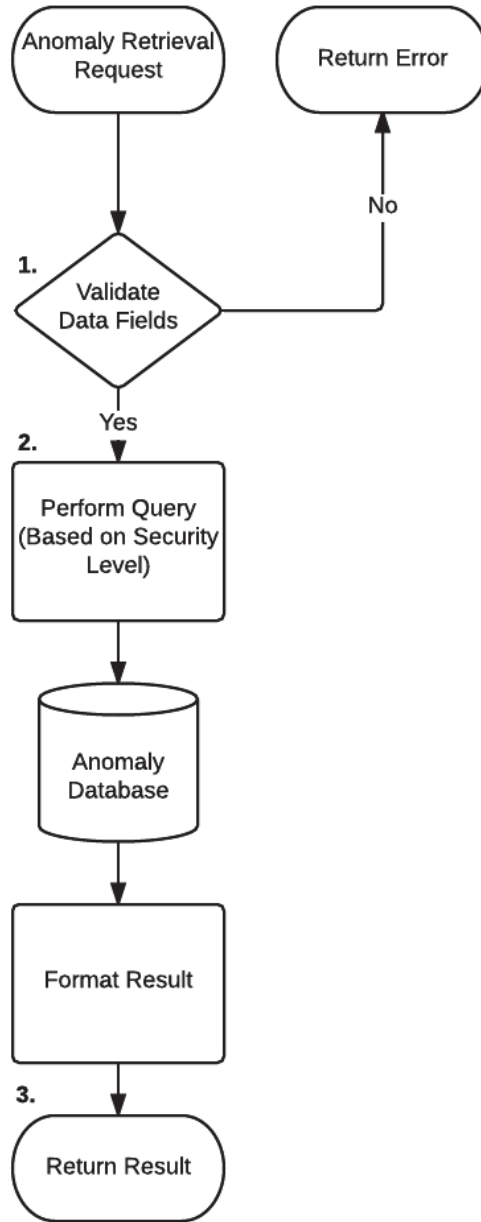


Figure 6.4: The flowchart for retrieval requests of stored anomaly reports. 1. Input validation ensures the request fields respect requirements. 2. The query is performed by filtering on record metadata with the appropriate security levels. 3. The result set is formatted and sent to the user.

6.2.3 Ship Profile Editing

Ship profiles aggregate all necessary information to identify a vessel. Additionally, they store references to all reported anomalies on the vessel in question. Providing operators with access to ship profiles is an important step in ensuring that vessel data is up to date and as accurate

as possible (see item 4, Figure 6.2). The goal is to allow users to correct misspelled ship names, incorrect MMSIs or other erroneous identifiers that may have sifted through to the database. These responsibilities entail, however, a user class with a higher order of access rights than regular operators as they have the potential to damage the data and the work of their peers. This user class is the Ruling Authority (see Section 5.3.4).

These users may edit vessel identity information, anomaly report information or rules on the legitimacy of a vessel's associated anomalies. The procedure for handling these requests is detailed in the list below (accompanied by Figure 6.5):

1. Verify that all input fields respect requirements. Produce field specific error messages otherwise. Also check that a ruling authority issued the query.
2. Find the requested ship in the database. Warn the operator if the ship doesn't exist.
3. Check the ruling authority's access rights to individual records affected by the query request. Warn the user if he has insufficient rights to perform the change.
4. Apply the changes to the ship profile.
5. Return a successful transaction message.

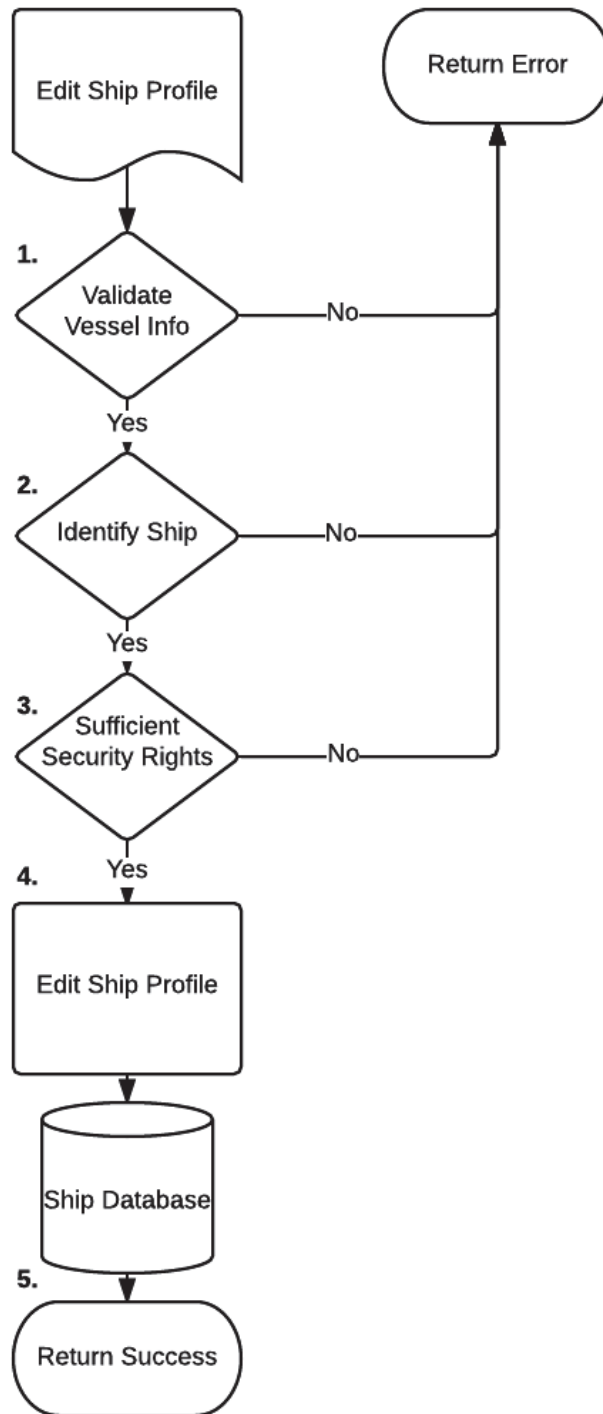


Figure 6.5: Ship profile editing flowchart 1. Validate that the query is issued by a ruling authority account and vessel information fields. 2. Search for the ship in the database. 3. Verify that the ruling authority has sufficient security access to edit the desired records. 4. Edit the ship profile and store the changes in the database. 5. Return a successful transaction message.

6.2.4 Ship Profile Retrieval

The ship profile retrieval allows operators to inspect all of SARA's amassed information on a given ship (see item 4, Figure 6.2). The flow of operations is nearly identical to that of ship profile editing requests. Figure 6.6 below, provides an overview of the major steps for the request.

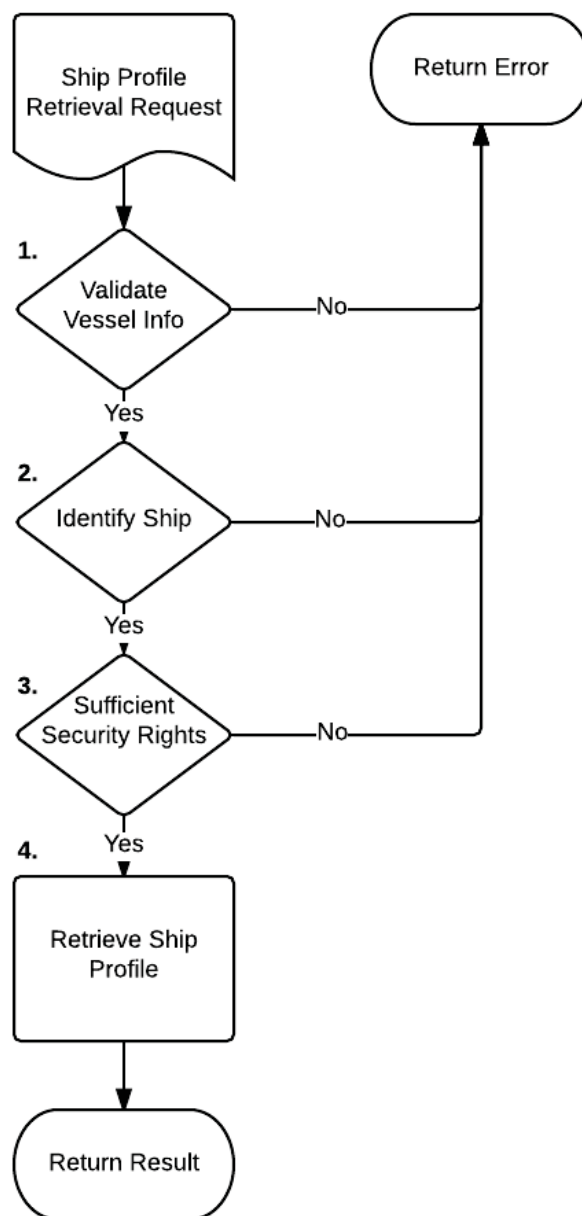


Figure 6.6: Ship profile retrieval flowchart. 1. Validate vessel information fields. 2. Search for the ship in the database. 3. Verify that the operator has sufficient security access to retrieve the desired records. 4. Retrieve the ship profile and return the result to the user.

6.2.5 User Creation and Management

New user accounts may only be added by administrators (See 5.3.4). These participants issue commands through the service interface, like all other user types, the main distinction being that their access rights allow them to perform a distinct set of actions reserved to user management. Certain record parts may require additional security access rights to edit.

Figure 6.7 below, presents the flowchart for user profile creation by an administrator.

1. Upon reception, SARA must validate that the query was sent by an administrator.
2. The new user information is verified to ensure that all required fields are met. The user information is also checked against the user database for existing users with the same name and handle, in order to avoid that an already existing user receives a second account.
3. The user account is created and stored in the database.
4. The user handle and a temporary password are transmitted to the user.

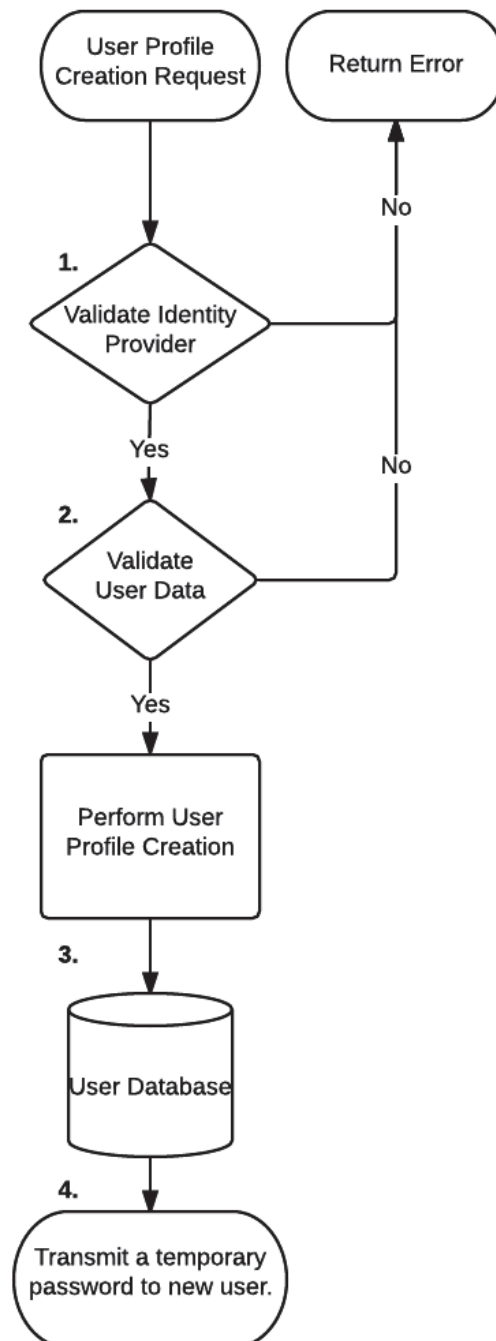


Figure 6.7: User creation flowchart. 1. SARA validates the query was sent by an administrator. 2. The new user information is validated for correctness and against potential existing accounts in SARA's user database. 3. The new user account is created and stored in the database. 4. The user handle and temporary password are sent to the user.

6.2.6 Authentication

In an operational context, SARA alternates between the role of Information Provider and Consumer as users either request anomalies stored in the system or submit them. From a security and services perspective, however, the roles are immutable with the user always occupying the role of Service Consumer and SARA occupying the role of Service Provider, as illustrated in figure 6.8.

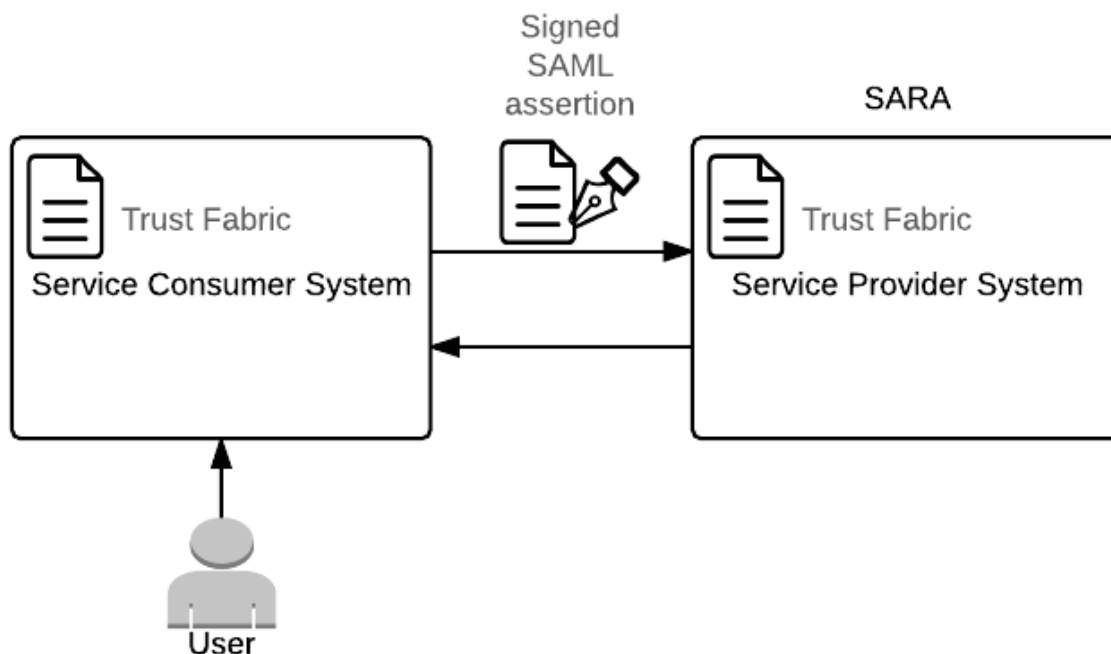


Figure 6.8: Service architecture in the context of NIEM. SARA acts as a service provider to a 3rd party application (the consumer system) through which the users (operators) interact. Both the consumer and provider maintain a copy of the trust fabric, a certificate issued by a NIEM certificate authority listing all members vetted for the network. Additionally, all communication between consumer and provider must be preceded by a SAML assertion, signed by the consumer.

Generally, users are expected to communicate to SARA through a third party application acting as a consumer system on their behalf. The user logs in through the consumer system, which then propagates requests to SARA. In NIEM architecture, the service consumer and service provider are trusted systems, entrusted to operate within a NIEM network. In terms of authentication and access management, each network may enforce its own standard. Security Assertion Markup Language (SAML) is used here as an example architecture already in use within MISE (see section 6.6 for information about MISE). SAML is a markup language standard for authorization built around single sign-on access control. It allows users to log in to multiple independent but related systems with a single account. This authorization scheme is therefore particularly useful, if SARA is integrated into a larger network. All trusted systems are issued a Trust Fabric document signed by the Certificate Authority (CA) in charge of the network. The trust fabric is an XML document in SAML metadata containing a list of all trusted systems validated to operate within the network.

It is renewed periodically and must be re-obtained by all trusted systems in order to continue operating within the network. The trust fabric forms the cryptographic foundation of trust within a network. The trust fabric is digitally signed by the CA with a private key and incorporates the public key in the certificate. The public key can be used to confirm the authenticity of the signature and therefore of the trust fabric.

The following list accompanied by the figure 6.9, showcase the typical sequence of events involved in a communication exchange between SARA and a consumer over a secure internet connection.

1. Signed certificates are required and exchanged at both ends of the connection client (consumer) and server (provider). Each side confirms the authenticity of the other's certificate thus establishing the Secure Socket Layer (SSL) connection.
2. Once the connection is established, the consumer POSTs user attributes to the provider by creating a SAML assertion. The assertion is digitally signed by the service consumer and can be used to lookup the identity of the trusted system in the trust fabric.
3. If the assertion is valid, the service provider stores the assertion in a session and returns a cookie response header to track the session and a status code of 200 (OK). If the assertion is invalid, the service provider returns a status code of 403 (Forbidden).
4. Since there is significant overhead in delivering SAML assertions and validating signatures, the document is sent only once and tracked as a session for subsequent service requests. This allows the system to remain RESTful for all transactions outside of authentication.
5. The session has a timed expiration as well as an inactivity expiration. However, the service consumer can POST to a logout URI, as a courtesy to the service provider, if it knows that interactions have terminated.

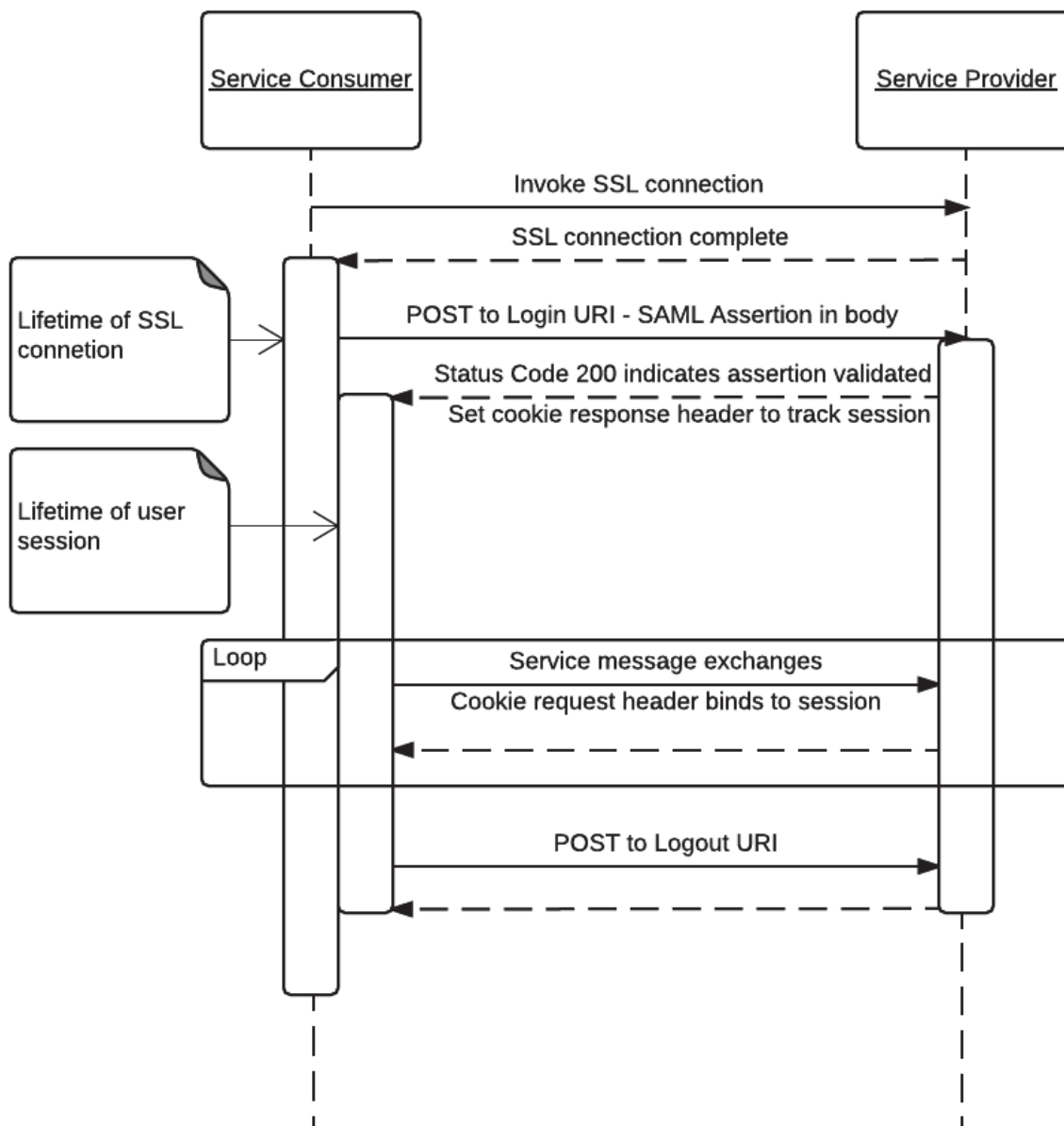


Figure 6.9: Sequence diagram of communication exchange. All communication between consumers and providers takes place over HTTP with a SSL which ensures the encryption of messages. User login information is transmitted over a SAML assertion signed by the service consumer, therefore ensuring the message's authenticity. The service provider creates user sessions if all the security information is validated and service messages can continue without any additional overhead. The service consumer can POST to a logout URI, as a courtesy to SARA, to terminate the user session and free up resources.

6.3 Database and Data Structure

One of the prerequisites for interoperability and efficient information discovery and sharing is to use a standard information format, in this case the NIEM. As such, to minimize development effort and ease the mapping between the NIEM and SARA underlying data structure, NIEM data structure should be followed as closely as possible when designing SARA's data structures. Also, relying on an existing and widely-adopted standard for object description reduces the time needed for design and implementation, as several resources already exists to support the needed development. In addition, the common vocabulary basis between the data structure and the NIEM exchange reports will simplify the field mapping for both the document readers and the developers that will have to implement SARA. The designed NIEM-based anomaly IEPD (see section 4.4) is composed of four main data blocks :

- The vessel;
- The anomaly;
- The contact information;
- The block metadata.

A vessel can be linked to multiple anomalies, where each anomaly will have a single point of contact. An anomaly report can only designate a single reporting entity, but a reporting entity may issue any number of anomaly reports on any number of vessels. Several blocks of record metadata could be defined for a single anomaly; however, this is not expected to occur in an operational context, so we will consider that the whole anomaly and all the information that it is made up of will be submitted in a single block of record metadata.

That data structure must also reflect the kind of queries that are likely to be made by the users of SARA. Referring to section 4.5, the data structure must accommodate efficient retrieval of anomalies based on:

- The vessel;
- The anomaly type;
- The geographical area of interest;
- The timestamp of the anomaly.

To this end, we will be leveraging database solutions offering a flexible schema that allow us to make changes quickly without losing important data that may have been acquired along the way.

One notable use case that influenced the decision behind the minimum amount of information required for an anomaly report to be relevant or useful, is the instance of a ship that exhibits anomalous behavior and yet can't be identified. This creates a bit of an oddity where a report aimed at highlighting the inappropriate behavior of a ship doesn't reference a ship. However, since

the time and place of the anomaly are recorded, it is believed that this information is still useful for future reference. By combining the report data with other sources of information, it would be possible to infer the identity of the unidentified ship at a later date and extract some value from the report.

With regards to the type of the NIEM element, mapping can easily be made between it and the corresponding database type. For instance, NIEM TextType translates to a String or Char sequence.

6.3.1 Vessel Profile Data Structure

Ship profiles are built over time using available AIS reports or other types of data from external sources. These profiles must be subjected to a lot of scrutiny from automated processes and trusted users alike, in order to establish reliable ground truth. These processes reside in dedicated software components and likely dedicated hardware in order to decouple their work from the rest of SARA's duties.

Anomaly reports also contribute to the ship profile when converted into a reliability rating for the ship in question. As anomaly reports are accumulated for each ship, the reliability rating is refined, eventually becoming a usable indicator for operators. Figure 6.10 presents the Unified Modeling Language (UML) design of the Vessel Profile data structure.

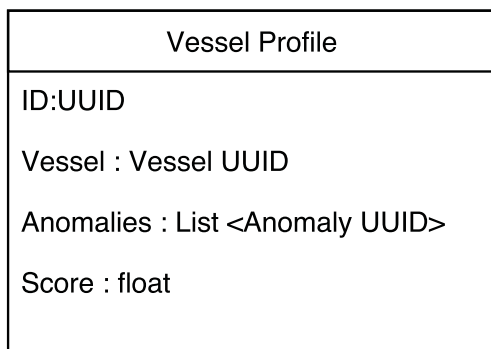


Figure 6.10: UML Design of the Vessel Profile data structure.

6.3.2 Vessel Data Structure

The ships' data structure is designed to help users create high quality anomaly reports. The goal is for the data describing a ship to be as exhaustive and robust as possible, so that users may confidently identify anomalies. Figures 6.11 and 6.12 present the NIEM-M fields available to describe a vessel within this framework.

VesselType		
anyAttribute		
VesselAugmentation	[0..1]	VesselAugmentationType
VesselActivityHistorySummaryText	[0..1]	TextType
VesselClassificationSocietyName	[0..1]	ClassificationSocietyNameType
VesselCargoOnBoardIndicator	[0..1]	boolean
VesselCertificateOfFinancialResponsibilityOperator	[0..1]	EntityType
VesselCVSSAOnBoardIndicator	[0..1]	boolean
VesselIncidentHistorySummaryText	[0..1]	TextType
VesselIceClassification	[0..1]	VesselIceClassificationType
VesselDoubleHullIndicator	[0..1]	boolean
VesselInterestHistoryList	[0..1]	VesselInterestHistoryListType
VesselLongshoreWorkIndicator	[0..1]	boolean
VesselNonTankVesselResponsePlanIdentification	[0..1]	IdentificationType
VesselNonTankVesselResponsePlanIndicator	[0..1]	boolean
VesselOtherRegistration	[0..*]	VesselOtherRegistrationType
VesselSubCategory	[0..1]	anyType
VesselViolationHistorySummaryText	[0..1]	TextType
VesselPropulsion	[0..1]	anyType

Figure 6.11: NIEM-M design of the vessel type.

VesselAugmentationType		
anyAttribute		
VesselBeamMeasure	[0..1]	MeasureType
VesselCallSignText	[0..1]	TextType
VesselCargoCategory	[0..1]	anyType
VesselCategory	[0..1]	anyType
VesselCDCCargoOnBoardIndicator	[0..1]	boolean
VesselCharterer	[0..1]	EntityType
VesselClass	[0..1]	anyType
VesselContactInformation	[0..1]	ContactInformationType
VesselCruiseSpeedMeasure	[0..1]	MeasureType
VesselDeckConfigurationText	[0..1]	TextType
VesselDescriptionText	[0..1]	TextType
VesselDOCCertificate	[0..1]	CertificateType
VesselDraftMeasure	[0..1]	MeasureType
VesselGrossTonnageValue	[0..1]	nonNegativeInteger
VesselHullNumberText	[0..1]	TextType
VesselIdentification	[0..*]	IdentificationType
VesselImage	[0..1]	ImageType
VesselIMONumberText	[0..1]	TextType
VesselISMCodeText	[0..1]	TextType
VesselISSC	[0..1]	InternationalShipSecurityCertificateType
VesselLocation	[0..1]	LocationType
VesselMaximumSpeedMeasure	[0..1]	MeasureType
VesselMMSIText	[0..1]	TextType
VesselName	[0..1]	ProperNameTextType
VesselNationalFlag	[0..1]	anyType
VesselNavigationStatus	[0..1]	StatusType
VesselOfficialCoastGuardNumberText	[0..1]	TextType
VesselOperationalConditionOfEquipment	[0..2]	anyType
VesselOperator	[0..1]	EntityType
VesselOverallLengthMeasure	[0..1]	MeasureType
VesselOwner	[0..1]	EntityType
VesselSafetyManagementCertificate	[0..1]	CertificateType
VesselSCONUMText	[0..1]	TextType

Figure 6.12: NIEM-M design of the vessel augmentation type.

As one can see, all the fields are optional. This enables one to select only those needed, while retaining the structure and naming to minimize vocabulary alignment and development work in translating between the database and messaging components of SARA. Figure 6.13 presents the

design of the Vessel data structure from its NIEM-M components.

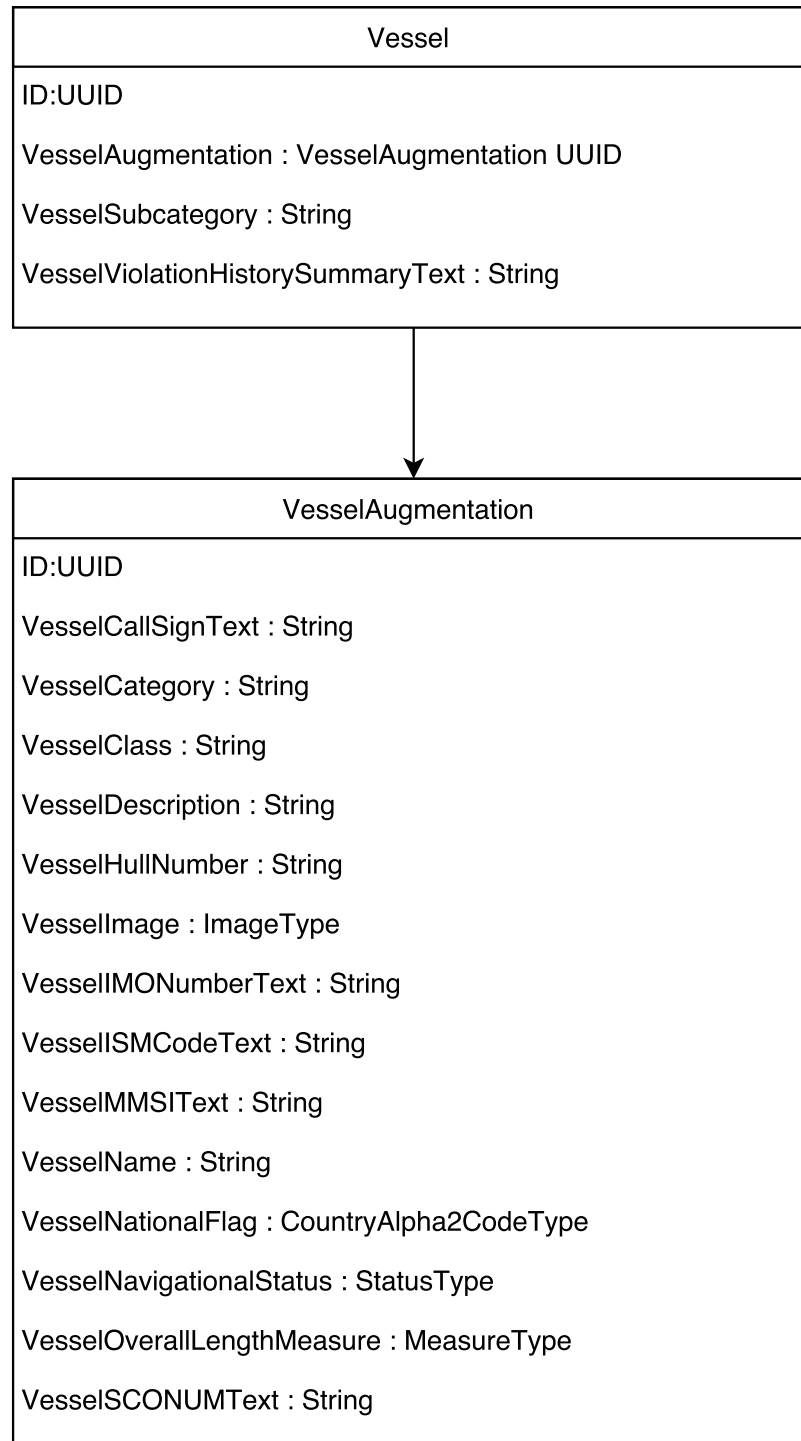


Figure 6.13: UML Design of the Vessel data structure.

6.3.3 Anomaly Data Structure

The anomaly data structure, shown in figure 6.14 can be reused as is. Note that the inclusion of ContactInformation and its ContactEntity will produce a very deep and nested view of the database.

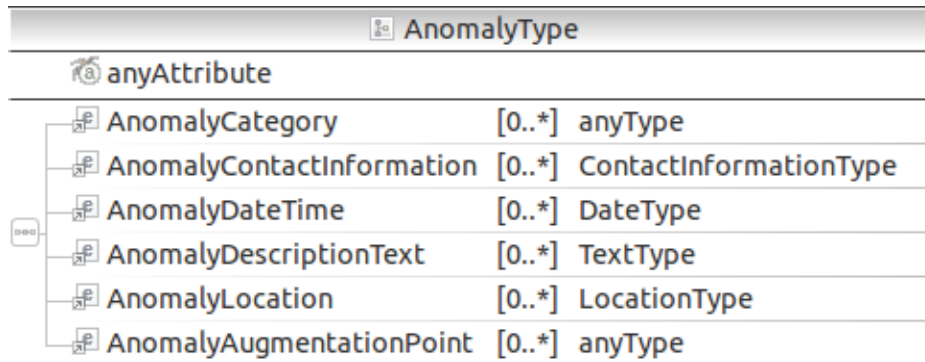


Figure 6.14: NIEM-M design of the anomaly.

Figure 6.15 presents the design of the Anomaly data structure from its NIEM-M components.

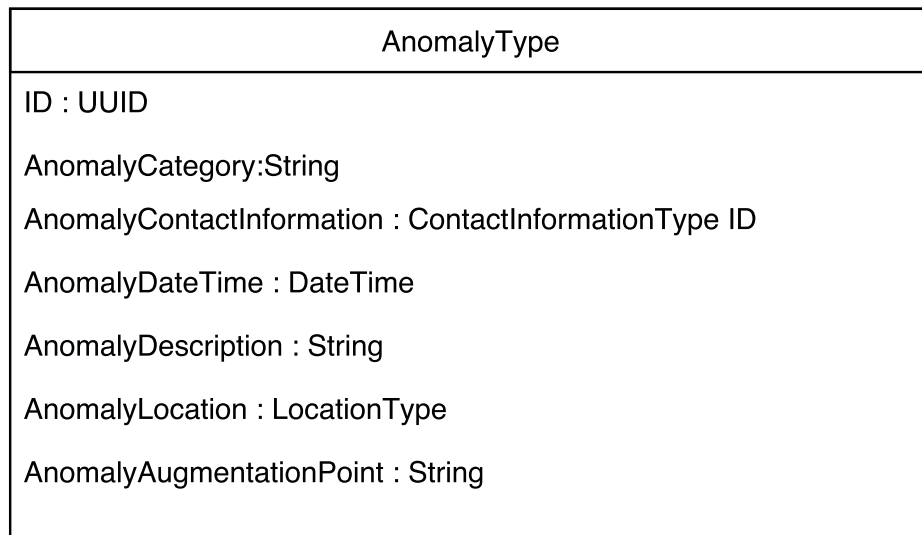


Figure 6.15: UML Design of the Anomaly data structure.

6.3.4 Contact Information Data Structure

This is the author of the anomaly. Human users are not the only ones able to author an anomaly report. In fact, it is expected that most reports will be generated by automated processes. From a

structural perspective, there is no difference between the two. In all cases, the required fields are taken from the NIEM Contact Information (figure 6.16) and its required EntityType. Note that the entity under ContactInformation can be both an organization (figure 6.17) or a person (figures 6.18 and 6.19), each with their respective specific data element.

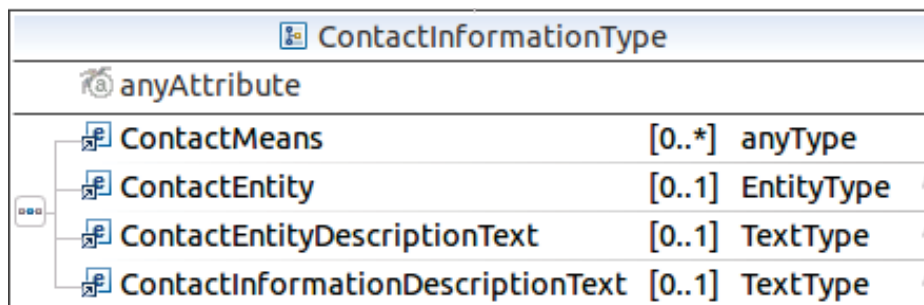


Figure 6.16: NIEM-M design of the contact information data structure.

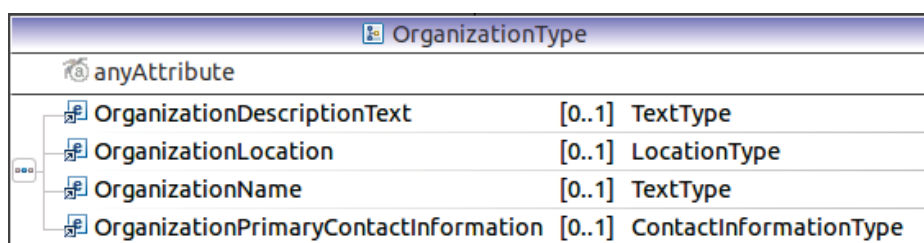


Figure 6.17: NIEM-M design of the organization data structure.

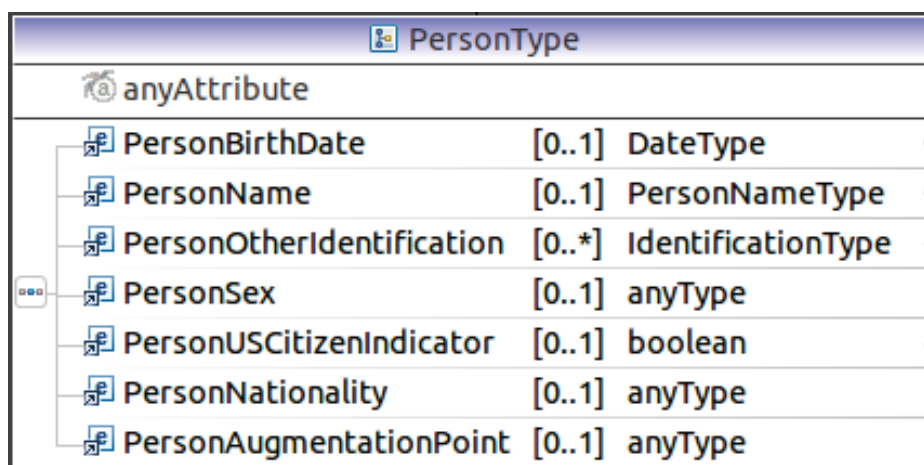


Figure 6.18: NIEM-M design of the person data structure.

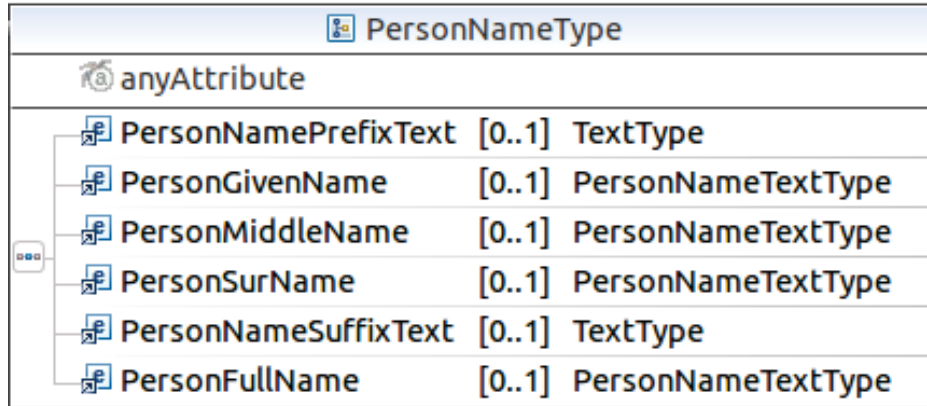


Figure 6.19: NIEM-M design of the person name data structure.

The ContactMeans can be a fax or phone number, an email or Uniform Resource Locator (URL), an address, and so on. This is a particularity to keep in mind when designing the actual database structure. Regarding the actual data fields, it would be advisable to associate them all with the Contact Information and Organization, limiting the Person and Person Name to only the name of the individual to contact and perhaps their nationality, unless specific requirements arise during the implementation. Figure 6.20 presents the design of the Contact Information data structure from its NIEM-Core components.

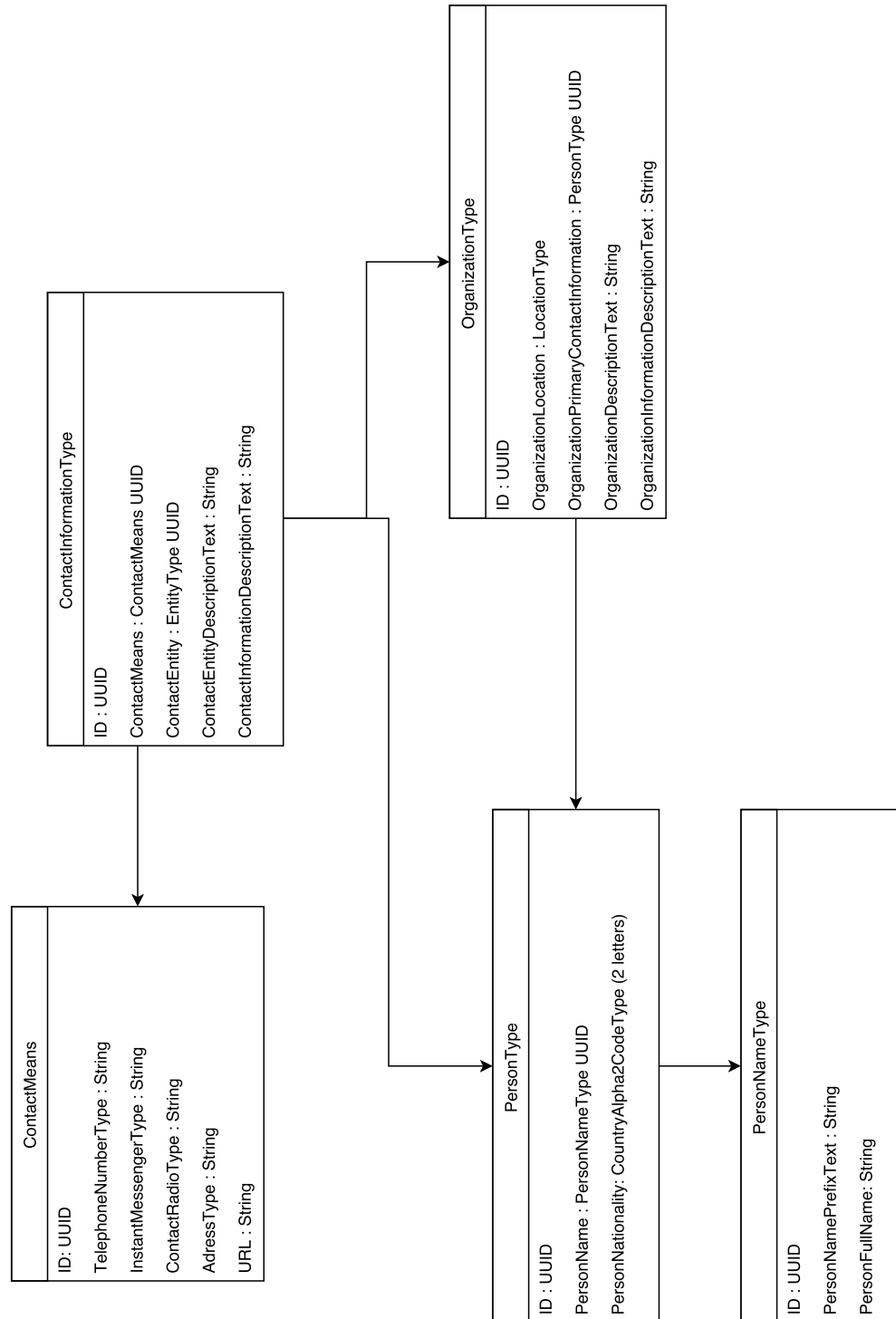


Figure 6.20: UML Design of the Contact Information data structure.

6.3.5 Block Metadata Data Structure

All the block metadata elements should be reproduced in the database, to ensure interoperability with other systems and compliant security markings and information entitlement.

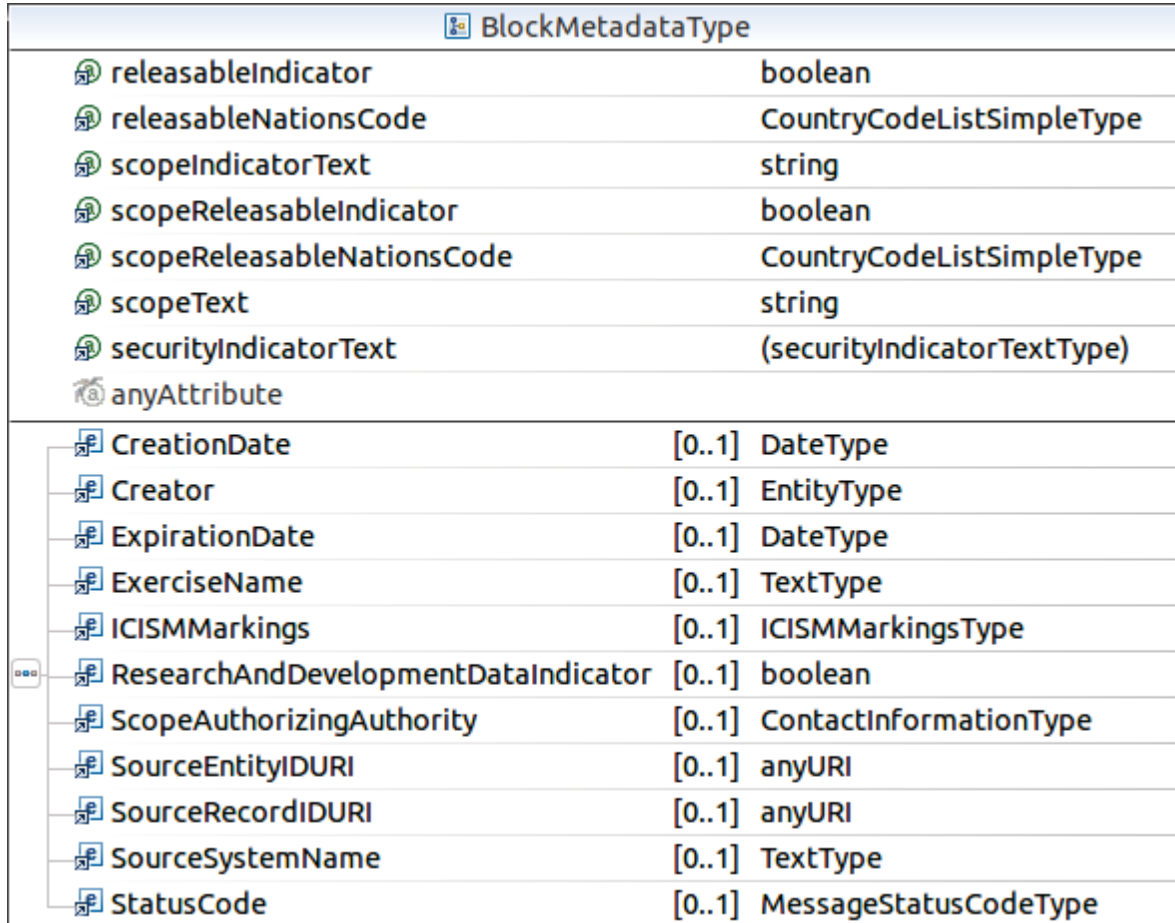


Figure 6.21: NIEM-M design of the block metadata.

6.3.6 Proposed Database Table Structure

The UML design of the proposed database structure is shown in figure 6.22. The fact that an Entity type can be either a Person or an Organization makes the design a little more complex. Also, to respect the NIEM schema, some of the type definitions have been omitted, but the reader should be aware the table structure can expand much more (to include MeasureType, LocationType, AddressType, etc.).

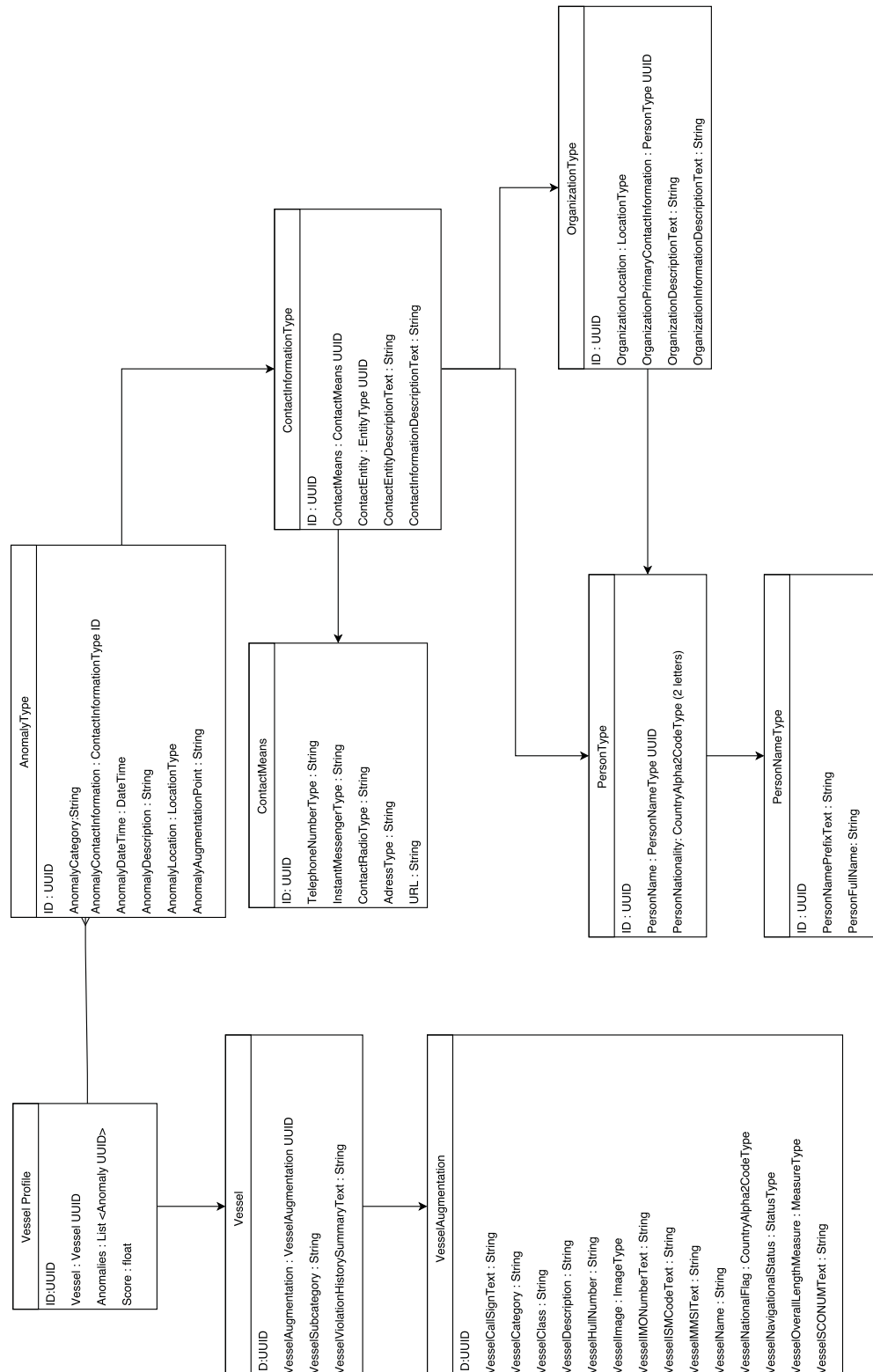


Figure 6.22: Database design in UML.

The NoSQL version, assuming a document-based database, would be very close to what is shown in Listing 6.1.

Listing 6.1: NoSQL document structure example

```

1 {
2   "UUID" : UUID,
3   "Anomaly": [
4     {
5       "AnomalyCategory" : string ,
6       "AnomalyDateTime" : {DateTime sturcture},
7       "AnomalyDescription" : string ,
8       "AnomalyAugmentationPoint" : string ,
9       "AnomalyLocation" : {},
10      "AnomalyContactInformation" : {
11        "ContactMeans" : {
12          "TelephoneNumberType" : {},
13          "InstantMessengerType" : {},
14          "ContactRadioType" : {},
15          "AdressType" : {},
16          "URL" : {}
17        },
18        "ContactEntity" : {
19          "OrganizationType" : {
20            "OrganizationLocation" : {LocationType},
21            "OrganizationDescriptionText" : string ,
22            "OrganizationInformationDescriptionText" : string ,
23            "OrganizationPrimaryContactInformation" : {
24              "ContactMeans" : {
25                "TelephoneNumberType" : {},
26                "InstantMessengerType" : {},
27                "ContactRadioType" : {},
28                "AdressType" : {},
29                "URL" : {}
30              },
31              "ContactEntity" : {
32                "PersonType" : {
33                  "PersonNationality" : string ,
34                  "PersonName" : {
35                    "PersonNamePrefix" : string ,
36                    "PersonFullName" : string
37                  }
38                }
39              "ContactEntyDescription": string ,
40              "ContactInformationDescriptionText" : string
41            }
42          }
43        }
44      },
45      "ContactEntyDescription" : string ,
46      "ContactInformationDescriptionText" : string
47    }
48  },
49  {Anomaly 2},
50  {Anomaly 3},
51  ...
52 ],

```

```

54  "Vessel": {
55    "VesselAugmentation":{
56      "VesselCallSignText" : string ,
57      "VesselCategory" : string ,
58      "VesselClass" : string ,
59      "VesselDescription" : string ,
60      "VesselHullNumber" : string ,
61      "VesselImage" : {ImageType fields},
62      "VesselIMONumberText" : string ,
63      "VesselISMCodeText" : string ,
64      "VesselMMSIText" : string ,
65      "VesselName" : string ,
66      "VesselNationalFlag" : string ,
67      "VesselNavigationalStatus" : {Navigational Status fields},
68      "VesselOverallLengthMeasure" : {MeasureTyp[e fields , eg. value and unit]},
69      "VesselSconumText" : string
70    },
71    "VesselSubcategory" : string ,
72    "VesselViolationHistorySummaryText" : string ,
73  },
74  "score": float
75 }

```

The blockmetadata elements have not been added to the UML and JSON representations, to avoid overcomplexity for the readers; however, the entitlement scheme should be applied and stored in the database and linked to the correct information pieces.

6.3.7 SQL Versus NoSQL

This section presents the difference between SQL and NoSQL databases with regards to the SARA application and its proposed relationship to the NIEM model. Table 6.1 presents an overview of SQL and NoSQL differences (from [56]).

	SQL	NoSQL
Data storage	Stored in a relational model, with rows and columns. Rows contain all of the information about one specific entry/entity, and columns are all the separate data points.	The term NoSQL encompasses a host of databases, each with a different data storage model. The main ones are: document, graph, key-value and columnar.
Schemas and Flexibility	Each record conforms to a fixed schema, meaning that the columns must be decided and locked before data entry and each row must contain data for each column. This can be amended, but it involves altering the whole database and going offline.	Schemas are dynamic. Information can be added on the fly, and each row (or equivalent) doesn't have to contain data for each column.

	SQL	NoSQL
Scalability	Scaling is vertical. In essence, more data means a bigger server, which can get very expensive. It is possible to scale an Relational Database Management System (RDBMS) across multiple servers, but this is a difficult and time-consuming process.	Scaling is horizontal, meaning across servers. These multiple servers can be cheap commodity hardware or cloud instances, making it a lot more cost-effective than vertical scaling. Many NoSQL technologies also distribute data across servers automatically.
Atomicity, Consistency, Isolation, Durability (ACID) Compliancy	The vast majority of relational databases are ACID compliant.	Varies between technologies, but many NoSQL solutions sacrifice ACID compliancy for performance and scalability

Table 6.1: Overview of SQL and NoSQL differences, from [56].

As shown in section 6.3.6, the Vessel Profile could *easily* be stored as either :

- A Relational Data Model: highly structured table organization with rigidly-defined data formats and record structure
- A document data model: collection of complex documents with arbitrary, nested data formats and varying record format.

As mentioned in table 6.1, a document database can easily be modified to support the evolution of the NIEM and maritime domain schemas and is easier and less expensive to grow. Another major advantage of a NoSQL database is the future release of NIEM is expected to include JSON specifications. Since a document database usually stores JSON data, there would not be any needs for conversion. Also, JSON is a natural data structure for web services, as it can easily be handled by language such as Javascript.

If other more complex analyses have to be made on the anomaly database, NoSQL technologies can easily support the MapReduce process, or any other process related to big data. These might be less straight forward with SQL technologies.

The NIEM XML structure is well suited to both types of database without transformation. Some SQL and NoSQL technologies support XML directly (SQL Server and MongoDB for instance), although one should not store data in XML directly, as it is principally a technology to move data between databases and applications. See figure 6.23 for an example of XML storage in SQL server and MongoDB.

SQL Server	MongoDB
<pre> 1 <Data 2 xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" 3 xmlns:xsd="http://www.w3.org/2001/XMLSchema"> 4 <Id /> 5 <Number>0</Number> 6 <DateTime>2007-10-03T00:00:00</DateTime> 7 <ValueA>32</ValueA> 8 <ValueB>88</ValueB> 9 <ValueC>aiWGsNkpviVpcJpMur</ValueC> 10 <ChildData> 11 <ChildData> 12 <Foo>961</Foo> 13 <Bar>Zf</Bar> 14 <Baz>286</Baz> 15 </ChildData> 16 <ChildData> 17 <Foo>925</Foo> 18 <Bar>OLXSU</Bar> 19 <Baz>371</Baz> 20 </ChildData> 21 <ChildData> 22 <Foo>601</Foo> 23 <Bar>DmKpUJJFAYI</Bar> 24 <Baz>16</Baz> 25 </ChildData> 26 <ChildData> 27 <Foo>817</Foo> 28 <Bar>PCX</Bar> 29 <Baz>491</Baz> 30 </ChildData> 31 </ChildData> 32 </Data> </pre>	<pre> 1 { 2 "_id" : ObjectId("523754b7ce6b612760b77e1d"), 3 "Number" : 16, 4 "DateTime" : ISODate("2009-07-05T04:00:00Z"), 5 "ValueA" : 80, 6 "ValueB" : 4, 7 "ValueC" : "FfduFz", 8 "ChildData" : 9 [{ 10 "Foo" : 130, 11 "Bar" : "MfzaJW", 12 "Baz" : 98 13 }, 14 { 15 "Foo" : 137, 16 "Bar" : "mpoXghBh", 17 "Baz" : 302 18 }, 19 { 20 "Foo" : 35, 21 "Bar" : "HVxG", 22 "Baz" : 535 23 }, 24 { 25 "Foo" : 296, 26 "Bar" : "qtM", 27 "Baz" : 735 28 } 29] </pre>

Figure 6.23: SQL Server versus MongoDB representation of an XML document.

Comparison was made in [57] for many operations of read and write. NoSQL database, in this case MongoDB, was out performing SQL technologies. Their main conclusion was that MongoDB relies on writing memory-mapped files to disk and returns as soon as the memory has been updated, but before the OS has finished syncing the memory-mapped file. Conversely, SQL Server returns results only after the data has been reliably logged, and as such, write operations are slower. For queries against unstructured data, query speeds were significantly faster for MongoDB. SQL Server's query performance for the XML data type was exceedingly slow, especially when performing queries against tables with large numbers of records.

In the end, there is not a definite answer to the type of database needed by the application. If SARA's data needs are changing rapidly, or it needs high throughput, or its data is growing fast and needs to be able to scale out quickly and efficiently, maybe NoSQL is a fit. But if the data isn't changing in structure and SARA is expecting to experience moderate, manageable growth, it needs may be best met by SQL technologies.

6.4 Data Processing

The data processing unit is in charge of all the complex computations required to support the database. It risks becoming a bottleneck for SARA's core services. Its decoupling from the application server reflects the need to maintain SARA as available as possible to operator requests. To be clear, these processes may well be housed in the domain logic of the application server but

they are expected to be so computationally costly as to require their own dedicated hardware. Conceptually, committing dedicated resources to these tasks, allows SARA to operate as intended with calculated overhead.

The following subsections detail each component of the data processing unit, illustrated in figure 6.24.

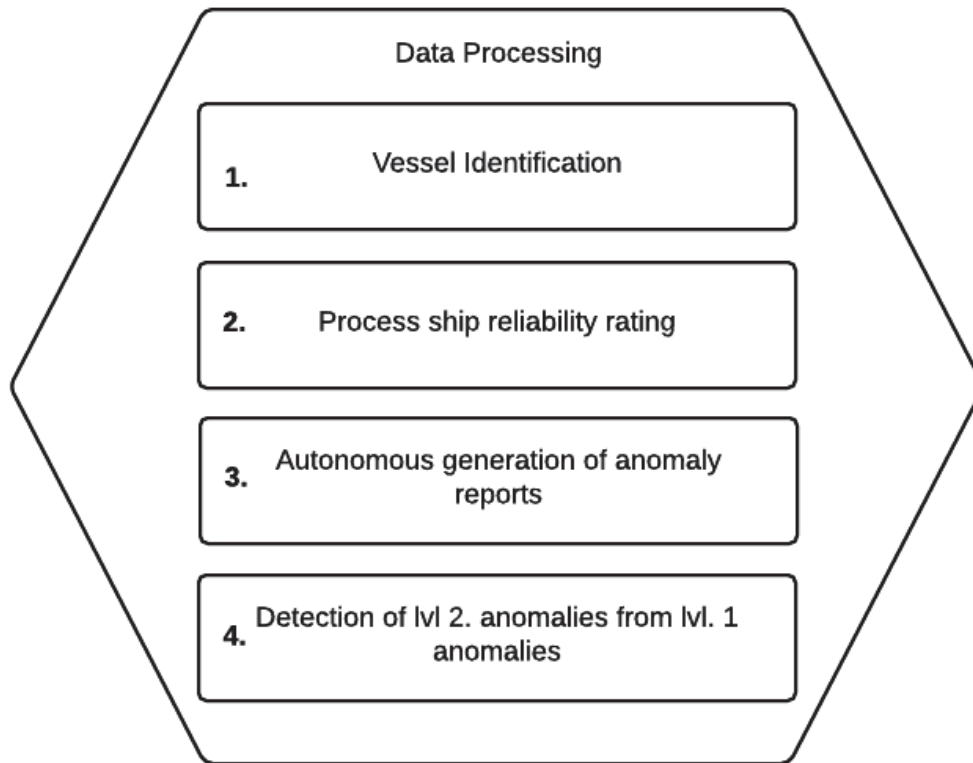


Figure 6.24: The Data Processing unit. 1. Since incoming anomalies may contain only partial information regarding the identification of a ship, establishing the relation between an anomaly and the vessel it references may not be straightforward. This process is therefore in charge of linking anomaly reports to the ship profiles internal to SARA. 2. A ship's reliability rating is a number that represents the accumulated product of all of the anomaly reports referencing the vessel. The rating is a representation of the severity of a ship's transgressions through its anomaly reports. The decoupling of the reliability rating from the application server is particularly useful when anomalies are being reported by an automated process. 3. Anomalies can be generated autonomously by a unit processing messages stored in an AIS database. 4. Level 1 anomalies are generated by automated processes analysing AIS reports. Level. 2 anomalies represent a deeper level of meaning requiring more processing power than the application server can afford to liberate.

6.4.1 Vessel Identification

Anomaly reports generally include the identification of a culprit vessel. SARA needs to identify this vessel within its own database in order to link it to the received anomaly report. However, there is an inherent ambiguity in reliably identifying a vessel from real-world data. Between incomplete, erroneous and malicious data, there are many ways in which any automated algorithm in charge of vessel identification may fail at its task. This component, illustrated in figure 6.25, encapsulates several means of dealing with the identification issue.

1. As a first step in identifying a vessel, all MMSI, vessel names and IMO numbers need to be separately analyzed for matches. If SARA cannot afford to respond to repeated search requests into its database, the subset of all MMSI, vessel names and IMO numbers may be cached locally for the identification process.
2. An approximate string matching algorithm such as the Baseza-Yates-Gonnet algorithm can be implemented for the search of vessel names, in order to account for name misspelling.
3. Lastly, an arbitration and ranking component is required to determine the order of the likeliest profile matches from the results of the search algorithm.

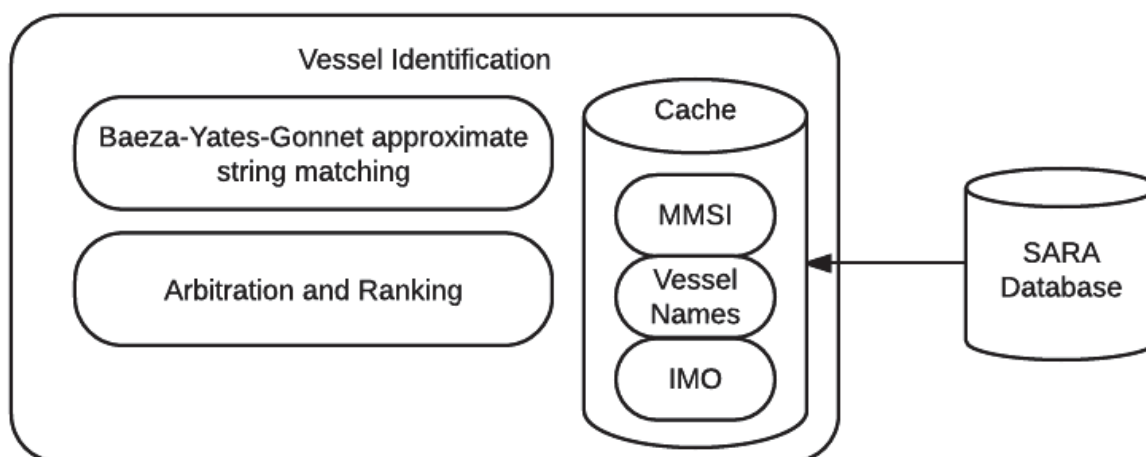


Figure 6.25: Subcomponents of vessel identification. A fuzzy search algorithm looks through existing vessel names, MMSI and IMO for similarities with a reported vessel. A local cache of data may mirror the vessel identification data stored in SARA's database, thus freeing up SARA to handle operator requests while the identification algorithm keeps working. A list of possible matching candidates is extracted and ordered by likelihood.

6.4.2 Reliability Rating

The reliability rating is the product of all anomaly reports associated to a vessel profile. It requires updating every time a new anomaly report is added to the database. The notion of liberating the

application server from the associated computational overhead is straightforward in this case. This decoupling is particularly relevant when automated processes are generating anomaly reports at a high rate.

6.4.3 Autonomous Generation of Anomaly Reports

When SARA is implemented, the AIS-related anomaly database will most likely be empty. Therefore, SARA requires a processing unit dedicated to the detection of anomalies from a large database. Another possibility is the use of a third-party system, TimeCaster or an exactEarth product for instance, to get AIS-related anomalies.

If the first option is chosen, the component should connect to an existing AIS repository, such as MSARI, in order to process anomaly reports from its vast amounts of recorded data. If MSARI is selected, a straightforward solution would be to use the data quality flags describing each AIS message already provided by MSARI. There are a total of nine data quality flags, including: position out of range, position not available, and invalid MMSI format. These quality flags could be mapped to the proposed taxonomy and transferred to the SARA database as level 1 anomalies. The range of anomalies would be limited, but still pertinent for SARA. Additional anomalies of level 1 could be detected by developing custom units processing MSARI AIS data.

6.4.4 Automatic Detection of Level 2 Anomalies from Level 1 Anomalies

As described in section 3, there is a link between anomalies of level 1 and of level 2. A level 2 anomaly is a history of AIS messages, attached to the same uniquely identified ship, describing a deviant pattern. Level 1 to level 2 transitions are captured in the taxonomy using an information quality approach.

For instance, a ship with a few anomalies of the “Error in AIS message - Voyage - Destination - Format not following the UN codes for ports and other locations - Not available” type (in other words, a ship with missing destination) will probably not be of great interest for an operator. However, if the same ship has a significant history of missing destinations, it might become interesting. To that effect, the taxonomy has a level 2 type corresponding to a history of missing destinations: “Incompleteness - Not defined - Voyage - Destination”.

As also discussed in that section, this concept of a transition between both levels opens the door to automatic detection of level 2 anomalies. If anomalies of type 1, i.e. errors in AIS messages, are reported (either by humans or systems) and stored in SARA database, it would be possible to create new anomalies of type 2 by automatically detecting patterns in these type 1 anomalies.

If we use the same example given above, a system could create an anomaly of type “Incompleteness - Not defined - Voyage - Destination” by sifting through the database and detecting history of having a predefined number of anomalies of type “Error in AIS message - Voyage - Destination - Format not following the UN codes for ports and other locations - Not available”. Such a system would use a set of rules to detect and create type 2 anomalies from type 1 anomalies. These exact rules should be carefully defined in close collaboration with end users. In addition to closely representing

the domain of knowledge, such collaboration would allow, a lower false alarm rate and hence better trust in the system.

A business rule management system, such as Drools, could be used to encode, match and act on such rules. This option is known to ease rules maintenance. Nonetheless because the rules are rather simple and without any apparent chaining, in-house code, optimised for the SARA database model, size and technology, could be sufficient. A trade-off analysis would be required, if such a system is developed.

6.5 Design Rationale

This section presents the rationale supporting the main design decisions presented above.

6.5.1 Service Architecture

The design choice with the most far-reaching consequences is the service-oriented architecture approach of SARA. The advantage of this architecture in terms of development is that we are treading on familiar ground. Although still a formidable undertaking, there is no shortage of guides and lessons learned in setting up a new web service. There is also an abundance of design patterns, applications and libraries readily available to cover any idiosyncrasy of SARA.

It is important, however, to properly understand the tradeoffs of this architecture with respect to its alternatives. Thinking of SARA as a web service implies that all its users would connect to it as a central and unique entity, despite the possibility that the actual location of the servers may be distributed over different locales. Users would connect to SARA much as they would to any other website and expect to see the same content no matter where they connected from. In that case, however, when the connection falters, the service becomes unusable. In an operational context where there is no available connection, SARA would have to rely on an external system to persist data locally so that it may be later unloaded into SARA when a connection becomes available.

In contrast, the architecture favored by the Coalition Shared Data Server (CSD) makes use of a distributed design in order to increase the robustness of its services. It operates in a context where an internet connection may prove unstable. The CSD allows far-flung clusters of units to coordinate when a connection is available and to operate independently when they become isolated. Each member of the CSD stores data on a local database and can share information with other members on request. The CSD architecture is much more complex than the one proposed to SARA and doesn't incorporate the notion of a singular reference database. It is, in fact, designed to service different interests and groups but retains the ability to share information when needed.

6.5.2 Query Management

The design decision to limit the query structure to a predefined list was reached in order to curb the logic complexity on the server and streamline communication with SARA. The alternative would imply the need for a complex query interpreter and query language.

Accordingly, should the user like to send a query that doesn't exist in SARA, the only way to obtain the desired result is to work with existing queries and manipulate the resulting data in an external analytic tool. This shifts some of the responsibility away from SARA and into the hands of the operator, when a complex analysis is required. But it assumes that the operator is both willing and able to perform the work.

Because of the large number of available attributes and all the possible ways to mix them into filters, the number of predefined queries required to cover all possibilities easily reaches the thousands. Developers can cover the most common cases involving attribute filters coupled with a simple logic (linked by AND logical operators). The most specific cases, which would involve OR or NOT logical operators for instance, could not be all covered. For the implementation phase, it is thus recommended to collaborate with operators to understand what kinds of queries are more pertinent for their work.

6.5.3 Authority Based User Creation

The choice to have new users added by an authority figure was adopted in view of enforcing a common style of identification of users and organizations. By limiting the creation rights to a select few, it's easier to ensure that a common and correct identification scheme is used in the process of creating new users. It would also reduce the chance of creating duplicate accounts. In short, this design choice is a measure of quality control.

It should be noted that there is little work involved tying SARA to this design choice and should the need arise to grant all potential users their own creation rights through the service interface, it could be easily implemented.

6.6 Integration into a NIEM Sharing Environment

By adopting the NIEM exchange model, SARA is leveraging existing NIEM networks to the fullest extent possible. As an example, the architecture plan for the MISE is shown in figure 6.26 below, with SARA depicted as a Trusted System.

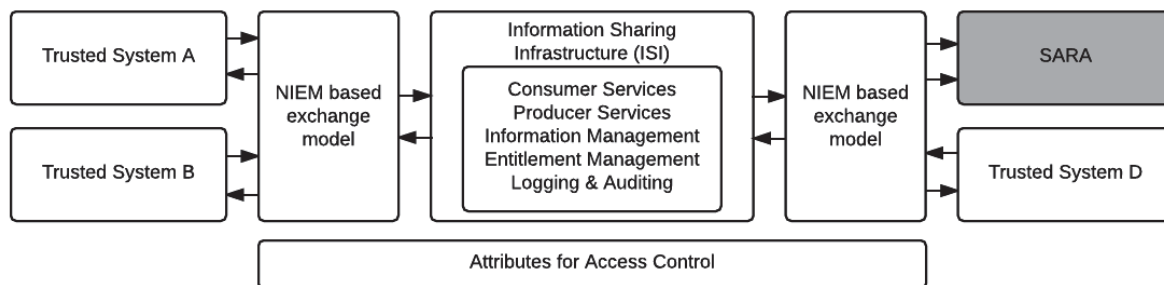


Figure 6.26: MISE: All member systems communicate over the internet with the ISI which provides services for acquisition and dissemination of data between trusted systems. All systems must communicate through the NIEM exchange model and specify the attributes for access control.

In MISE, the Information Sharing Infrastructure (ISI) acts as a central hub for information distribution between trusted systems. It simplifies the sharing mechanisms for all participants by offering a single framework linking all members. Trusted systems can only share information through the ISI, which regulates access rights on each shared product (entitlement management).

All trusted systems communicate over the open internet and can act equally as providers or consumers of data, so long as they communicate through the NIEM exchange model and adopt the ISI attributes for access control. Through these attributes, providers can specify sharing restrictions defined by their respective authorities and regulations.

Through MISE, SARA would then be able to share its anomalies and ship profiles over the ISI, generate new anomalies autonomously from data obtained on other trusted systems or directly ingest anomalies from users connected to another trusted system.

This page is intentionally left blank.

Part 7

Risks

This section presents the main risks associated with the development of SARA and is organized as follows:

- Section 7.1 presents the risks of handling ambiguities in ship identity.
- Section 7.2 discusses the risk in creating vessel profiles from anomaly reports.
- Section 7.3 presents the risk in the large volume of level 1 anomalies.
- Section 7.4 presents the risk of the lack of integration options for the anomaly retrieving components.

7.1 Handling Ambiguities in Ship Identity

An anomaly report submission contains information identifying the perpetrator vessel. Upon reception, SARA is expected to attach the anomaly to the ship profile described in the report. This is the process by which SARA adds value to operators. However, there is an inherent difficulty in uniquely identifying a vessel. For starters, the data identifying a vessel may be incomplete and lead to multiple matching candidates. When there are multiple possible matches, each could be ranked by the probability that it is the true match. Such a process would require considerable development effort and fine-tuning for it to be reliable. Another issue is to determine what degree of certainty is required in order for the association between vessel and anomaly to take place.

Additionally, it could turn out that the process of association is too computationally intensive, becoming a bottleneck for the rest of the system. In this case it is envisaged that the process could be decoupled from the regular flow of operations and instead, be scheduled in a periodic process on the data processing unit.

7.2 Creating Vessel Profiles from Anomaly Reports

Misspelled ship names are a common occurrence in AIS reports, as are incorrect MMSI, IMO, etc. Therefore, the automated creation of ship profiles from anomaly reports introduces the possibility of creating duplicate ship profiles from slightly different, erroneous or spoofed source data originating from the same ship. The issue is similar to the identification of vessels in that source data is unreliable by nature. In this case, the outcome is the creation of vessel profiles that have no bearing on reality, and end up polluting the database.

This issue may be further exacerbated by a process generating automated anomaly reports at very high frequency.

Initially, it may be advisable to exclude the option for automated ship profile creation. New ships would instead be added from a trusted database until a sizable number of profiles are amassed, thus reducing the occurrence of new ship profile creation once re-enabled.

7.3 Large Volume of Type 1 Anomalies

It was observed during development and maintenance of MSARI, see [58], that between 57 and 90 percent of AIS reports (variation depends on AIS sources and type: MSSIS, exactEarth, ...) have incomplete attributes (e.g. destination not available) and up to 3 percent had attributes out of range (e.g. heading greater than 360). Both these types of quality issues are in fact level 1 AIS-related anomalies. Considering that MSARI parses and stores about 66 million reports per day (see [59]), about up to 59 million level 1 anomalies could be extracted from MSARI per day. If all these anomalies are to be stored in SARA, we can expect serious pressure on its software and hardware. Therefore, it will have to be decided at the beginning of the development if automatic reporting of level 1 anomalies (errors in AIS messages) from an AIS database (e.g. MSARI) will be performed.

If yes, requirements will have to be derived from the anticipated high volume of anomalies to be stored and queried from SARA's database. Impacts on the system will be multiple: hardware selection and optimization, design and optimization of the physical data model, as well as the anomaly submission and retrieval components. However, lessons learned from MSARI development can be leveraged for this challenge. On the other hand, if there is no automatic reporting of level 1 anomalies from an AIS database, this would limit the capability of automatically detecting level 2 anomalies from level 1 anomalies, as discussed in section 6.4.4.

7.4 Integration of the Anomaly Retrieval Component

While section 2 proposes potential systems for the integration of the anomaly retrieval component, the integration challenge remains important. Indeed, a lot of constraints are imposed for ECDIS on board of Royal Navy ships: colors, shapes, etc. Therefore, integrating SARA's anomaly retrieval component to such ECDIS (e.g. SHINNADS or IMIC3) will come with non-negligible overhead

that may repel potential collaborators.

This page is intentionally left blank.

Part 8

Demonstration Application

An application was produced to help illustrate the value of sharing information about AIS-related anomalies.

The application was not built following the design recommendations made in this document. Instead, it was developed quickly to help market the overall idea further to the maritime security and safety community.

The demonstration focuses on the anomaly retrieving side of SARA, instead of on the anomaly reporting side. This decision was driven by the desire to advertise the added value of SARA by offering a focused story illustrating how it could be used by an operator concerned with maritime security and safety.

To be effective, the story had to be:

- realistic so that the potential users and decision makers are concerned,
- a showcase of the added value of SARA and
- simple.

This section is organised as follows :

- Section 8.1 presents the proposed scenario for the demonstration.
- Section 8.2 provides the detailed steps of the demonstration.

The Annex B provides the installation and launch instructions as well as troubleshooting.

8.1 Scenario

The demonstration is intended for maritime security and safety personnel on board for patrol.

Maritime Coastal Defence Vessel Project (MCDV) HMCS Goose Bay is on a patrol mission. Bridge watch keepers are looking at ships showing up on the radar. They are looking for Vessel Of Interest (VOI) for inquiry or closer inspection.

Two hours into that mission, four vessels are encountered with three of them having a history of AIS-related anomalies. Anomalies histories are provided by SARA.

End user : Bridge watch keeper on MCDV vessels.

Role : Drive the daily schedule of the ship, which depends on the ship's mission (e.g. missile shoot).

Mission : Patrol to identify other vessels.

Objective : Identify VOI for inquiry or closer inspection.

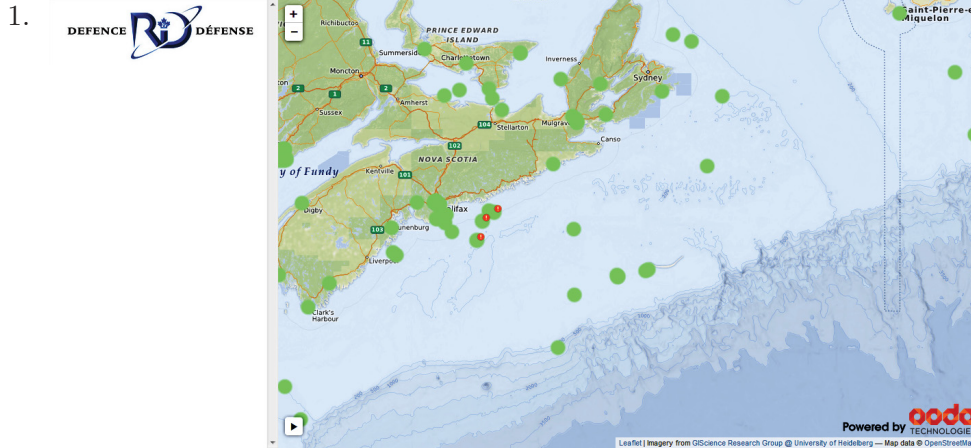
Workflow for the patrol mission :

1. When a vessel shows up on the radar,
2. the operator requests the anomaly history for this vessel by clicking on the contact.

The SARA query filters are set for one-year period, cover the Atlantic Sea and anomalies are provided from all partners.

8.2 Detailed Story

For each step of the scenario, a description of what appears on screen (*On screen*) and the actions to take (*Actions*) are provided.



Start the application.

On screen : Vessels in the bounding box: (min lon, min lat, max lon, max lat) =
(-67.39014, 40.88029, -51.56982, 48.22467)

Actions :

- (a) Explanation of what is SARA.
- (b) Explanation of the scenario.

2. Press *play* (bottom-left part of the screen)

On screen : Zoom on the MCDV contact.

- (a) The line behind the contact is its track since 10h00.
- (b) The dotted circle around the contact represents the radar range.
- (c) Everything that falls in that circle is of interest for the operator.

Actions : Explanation of what is on the screen (see *On screen* above). There is about 5 seconds to do so before next step.

3. **Vessel**

COLBY PERCE - 31603090
MMIS: 31603090
MNO: 8296169
Call sign: VCMF
Type: Fishing - (30)

Anomalies

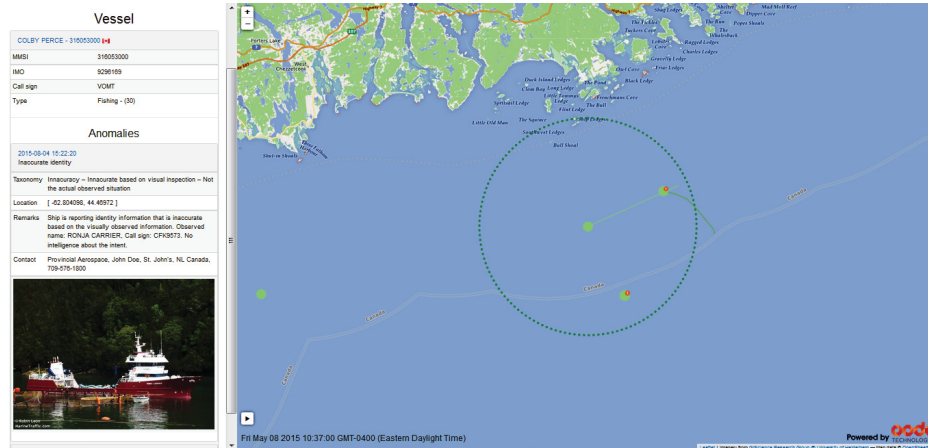
2015-05-04 10:02:23
Inaccurate identity

Taxonomy: Inaccuracy - Inaccuracy based on visual inspection - Not the actual observed situation

Location: [42.304396, 44.40972]

Remarks: Ship is reporting identity information that is inaccurate based on the visually observed information. Observed name: RONJA CARRIER, Call sign: CF93073. No intelligence about the vessel.

Contact: Provincial Aerospace, John Doe, St. John's, NL, Canada, 709-575-1800



Press *pause* for the first time (at around 10h37).

On screen : 2 contacts inside the radar range.

Actions :

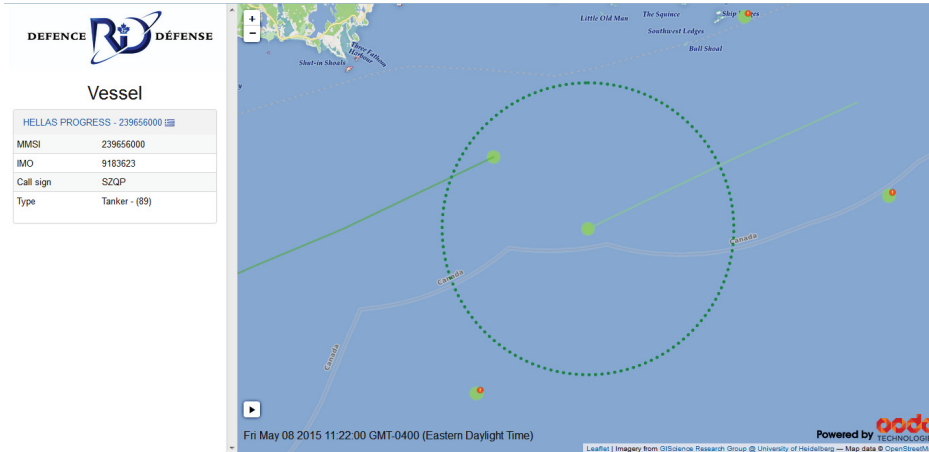
- Click on the MCDV contact. The vessel profile appears on the left side.
- Click on the contact in the right part of the radar range (vessel COLBY PERCE): 2 anomalies are recorded (displayed on the bottom left part of the screen).
- Click on the contact in the bottom part of the radar range (vessel STEPHANIE DANN): 1 anomaly recorded.

4. Press *play* for the second time.

On screen : MCDV is going forward, the previous 2 vessels are leaving the radar range and a third vessel is approaching.

Actions : There is a time gap of about 9 seconds until the next pause that can be used to provide further explanations about SARA and its added values.

5.



Press *pause* for the second time (at around 11h22).

On screen : 1 contact on the top part of the radar range.

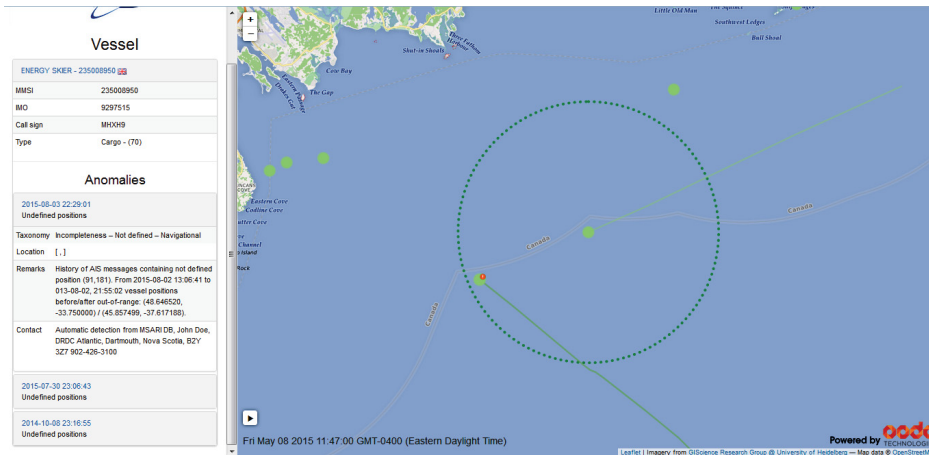
Actions : Click on the contact (vessel HELLAS PROGRESS): no anomaly recorded.

6. Press *play* for the third time.

On screen : MCDV is going forward, HELLAS PROGRESS is leaving the radar range and a fourth vessel is approaching.

Actions : There is a time gap of about 5 seconds until the next pause.

7.



Press *pause* for the third time (around 11h47).

On screen : 1 contact on the left part of the radar range.

Actions : Click on the contact (vessel ENERGY SKIER): 3 anomalies (same type) recorded.

This page is intentionally left blank.

Part 9

Conclusion

This document presented the concept of a system called SARA that shares AIS-related anomalies, called SARA. While it is a known fact that there are quality issues with AIS messages, very few of these AIS-related anomalies are persisted and exploited. SARA provides a means to share, persist, expose and use these anomalies to improve the maritime awareness.

This document, in addition to presenting the concept of SARA, proposed investigations on meta-data, based on the NIEM framework, to share the AIS-related anomalies as well as on a taxonomy to structure them. It was found that the taxonomy proposed fits well into the existing NIEM models; however, the taxonomy will have to be further adapted to fit end-user vocabulary and workflow. The resulting rework would not have a significant impact on the NIEM.

The document also proposed a design for SARA's future implementation. It was found that the implementation of such system implies few technical risks and can be done relying on mature technologies and architectures, such as service-oriented architecture, REST and SAML. However, half of the identified risks are related to the unique identification of a ship. This issue should be carefully considered before starting the implementation and the different options to uniquely identify a ship will have to be investigated.

Finally, it was deemed important to integrate SARA with existing AIS-related systems for multiple reasons: better integration with end-user workflow, benefit from collaboration between respected players in the maritime security, reduction of duplication of efforts related to that topic, etc. During this contract, some of the players of the maritime security were contacted and collaborations initiated. The project has raised the enthusiasm of the majority of the contacted federal and commercial parties. A workshop on the topic of AIS-related anomalies at DRDC-Atlantic would continue building these collaborations and share knowledge.

This page is intentionally left blank.

Bibliography

- [1] OSI Maritime. Ecpins. URL <http://osimaritime.com/solutions/ecpins-plus/ecpins>.
- [2] Offshore Systems Ltd. Electronic chart precise integrated navigation system (ecpins-w sub), tactical dived navigation. URL http://www.naval-technology.com/downloads/whitepapers/data_management/1808/.
- [3] Special Issue Altay News. Warship automatic identification system (w-ais), August 2011. URL http://www.altay.com.tr/PageGalleryFiles/PdfFiles/pdfen/2011/110812_Altay_News_WAIS.pdf.
- [4] MarineLink.com. Osi: Helping to chart the future of navigation. URL <http://www.marinelink.com/article/maritime-standards/helping-chart-future-navigation-864>.
- [5] James Day. Integrated picture, integrated response. Vanguard, January 2009. URL <http://www.vanguardcanada.com/2009/02/01/integrated-picture-integrated-response/>.
- [6] Thales. Commander c3. URL <https://www.thalesgroup.com/fr/worldwide/defense/commander-c3>.
- [7] Ian Coutts. Commander c3: Enhancing situational awareness, January 2013. URL <http://www.vanguardcanada.com/2013/01/30/commander-c3-enhancing-situational-awareness/>.
- [8] Thales Canada. Commander c3, maritime mission management system. URL https://www.thalesgroup.com/sites/default/files/asset/document/c3_datasheet.pdf.
- [9] Office of the Assistant Secretary of the Navy for Research. Gccs-m global command and control system - maritime. URL <http://www.secnav.navy.mil/rda/Pages/Programs/GCCSM.aspx>.
- [10] Andrew Wareing. Port of halifax gets puretech surveillance software, May 2007. URL <http://www.canadiansecuritymag.com/news/transportation/port-of-halifax-gets-puretech-surveillance-software-705>.
- [11] Ultra Electronics. Ultra awarded command & control system for the port of halifax, nova scotia, canada, April 2007. URL https://www.ultra-electronics.com/uploads/PressRelease/HPACCS_Press_Release_080507.pdf.

- [12] Provincial Aerospace. Mpa: A force multiplier. URL <http://www.provincialaerospace.com/LinkClick.aspx?fileticket=owxEeLFgkeY%3D&tabid=94>.
- [13] Maerospace. System and method for tracking and forecasting the positions of marine vessels, 2015. URL <https://patentscope.wipo.int/search/en/detail.jsf?docId=W02015127540&recNum=1&maxRec=&office=&prevFilter=&sortOption=&queryString=&tab=PCT+Biblio>.
- [14] Maerospace. Timecaster brochure, 2014. URL <http://maerospace.com/wp-content/uploads/2014/05/Maerospace-Brochure.pdf>.
- [15] Maerospace. Maerospace announces canadian dnd contract, July 2015. URL <http://maerospace.com/maerospace-announces-canadian-dnd-contract/>.
- [16] Maerospace Eric Meger. Technology challenges in maritime domain awareness & timecaste solutions. In *Canadian Fusion and Tracking Group*, November 2015.
- [17] exactEarth. Preliminary long form prospectus - english. Sedar, July 2015. URL <http://sedar.com/DisplayCompanyDocuments.do?lang=EN&issuerNo=00037341>.
- [18] exactEarth. Methods and systems for consistency checking and anomaly detection in automatic identification system signal data, 2015. URL <http://www.google.com/patents/US9015567>.
- [19] Greenline Systems. Greenline vessel selection system (vss), 2013. URL <http://www.greenlinesystems.com/vessel-risk-targeting/>.
- [20] Public Works and Government Services Canada. Greenline systems - sw integration (w8474-14at35/a), August 2013.
- [21] Paul Kerstanski. A maritime security example of how to risk assess cargo and crew. Greenline Systems, 2011. URL <http://www.greenlinesystems.com/2011/11/07/a-maritime-security-example-of-how-to-risk-access-cargo-and-crew/>.
- [22] Volpe. Tracking vessels on the saint lawrence seaway, 2002. URL <http://www.volpe.dot.gov/our-work/infrastructure-systems-and-technology/tracking-vessels-saint-lawrence-seaway>.
- [23] Volpe. M.s.s.i.s., 2012. URL <https://mssis.volpe.dot.gov/Main/>.
- [24] Volpe. Tv32 overview, 2012. URL <https://mssis.volpe.dot.gov/Main/tv32.php>.
- [25] *Global MDA Conference, 2nd Western Hemisphere MDA Workshop*, February 2009.
- [26] European Commission. Legal aspects of maritime monitoring & surveillance data, summary report. Technical report, DG Maritime Affairs & Fisheries, October 2008. URL http://ec.europa.eu/maritimeaffairs/documentation/studies/documents/legal_aspects_maritime_monitoring_summary_en.pdf.
- [27] Michael Davenport. Maritime anomaly detection workshop report and analysis. Technical Report CR 2008-275, DRDC CORA, March 2008.

- [28] Abbas Harati-Mokhtari, Alan Wall, Philip Brooks, and Jin Wang. Automatic identification system (ais): data reliability and human error implications. *Journal of navigation*, 60(03): 373–389, 2007.
- [29] Marco Balduzzi, Alessandro Pasta, and Kyle Wilhoit. A security evaluation of ais automated identification system. In *Proceedings of the 30th Annual Computer Security Applications Conference*, pages 436–445. ACM, 2014.
- [30] A Felski and K Jaskólski. The integrity of information received by means of ais during anti-collision manoeuvring. *TransNav: International Journal on Marine Navigation and Safety of Sea Transportation*, 7(1), 2013.
- [31] Clément Iphar, Aldo Napoli, and Cyril Ray. Detection of false ais messages for the improvement of maritime situational awareness. In *Oceans’ 2015*, 2015.
- [32] Walter L Perry, David Signori, E John Jr, et al. *Exploring information superiority: a methodology for measuring the quality of information and its impact on shared awareness*. Rand Corporation, 2004.
- [33] Eric S. Raymond. Aivdm/aivdo protocol decoding, May 2015. URL <http://catb.org/gpsd/AIVDM.html>.
- [34] N Bailey. Training, technology and ais: looking beyond the box. In *Proceedings of the Seafarers International Research Centre’s, 4th International Symposium Cardiff University*, volume 108, page 128, 2005.
- [35] Mike Davenport. Kbad final report. Technical Report CR 2008-002, DRDC CORA, April 2008.
- [36] Public Safety Canada. National information exchange model (niem), December 2015. URL <http://www.publicsafety.gc.ca/cnt/bt/niem/index-en.aspx>.
- [37] NIEM. National information exchange model - history. <https://www.niem.gov/aboutniem/Pages/history.aspx>, 2014. URL <https://www.niem.gov/aboutniem/Pages/history.aspx>. [Online; accessed November-2014].
- [38] Jeskell Inc. Niem and ucore content analysis and tagging proof of concept. http://www.jeskell.com/download/JSK_analytics_wp.pdf, August 2009. URL http://www.jeskell.com/download/JSK_analytics_wp.pdf. [Online; accessed November-2014].
- [39] Wikipedia. National information exchange model. http://en.wikipedia.org/wiki/National_Information_Exchange_Model, 2013. URL http://en.wikipedia.org/wiki/National_Information_Exchange_Model. [Online; accessed December-2013].
- [40] NIEM. Niem - webinars. <https://www.niem.gov/training/Pages/webinars.aspx>, 2014. URL <https://www.niem.gov/training/Pages/webinars.aspx>. [Online; accessed November-2014].

- [41] NIEM. Niem - tools catalog. <https://www.niem.gov/tools-catalog/Pages/tools.aspx>, 2014. URL <https://www.niem.gov/tools-catalog/Pages/tools.aspx>. [Online; accessed November-2014].
- [42] NIEM. Niem - online. <https://www.niem.gov/training/Pages/online.aspx>, 2014. URL <https://www.niem.gov/training/Pages/online.aspx>. [Online; accessed November-2014].
- [43] Scott Renner. A comparison of cursor-on-target, ucore, and niem. Technical report, The MITRE Corporation, 2012. URL http://www.mitre.org/sites/default/files/pdf/13_1152.pdf. [Online; accessed November-2014].
- [44] NIEM. National information exchange model - maritime. <https://www.niem.gov/communities/maritime/Pages/about-maritime.aspx>, 2014. URL <https://www.niem.gov/communities/maritime/Pages/about-maritime.aspx>. [Online; accessed November-2014].
- [45] Departement of Defense. Adoption of the national information exchange model within the department of defense. Memorandum, March 2013.
- [46] MISE. Niem-m data standards. <https://mise.mda.gov/drupal/node/42>, 2014. URL <https://mise.mda.gov/drupal/node/42>. [Online; accessed November-2014].
- [47] MISE. The national maritime domain awareness architecture plan. https://mise.mda.gov/drupal/sites/default/files/MDA_Arch_Plan_V_2.0_Release2_Full_1.pdf, November 2013. [Online; accessed November-2014].
- [48] Maritime Information Sharing Environment. National information exchange model - maritime artifact versioning plan. Technical report, NIEM, November 2014.
- [49] NIEM. How niem uses xml. URL https://www.niem.gov/training/Documents/Mod10_NIEM_PI_How_NIEM_uses_XML.pdf.
- [50] The White House. National strategy for information sharing and safeguarding. Technical report, 2012. URL https://www.whitehouse.gov/sites/default/files/docs/2012sharingstrategy_1.pdf.
- [51] Maritime Information Sharing Environment. Entitlement marking for iepd instances, 2014. URL <https://mise.mda.gov/drupal/node/126>.
- [52] Maritime Information Sharing Environment. Search parameters for notice of arrival, 2014. URL <https://mise.mda.gov/drupal/node/49>.
- [53] NIEM. Suspicious activity report iepd details, 2008. URL <https://tools.niem.gov/niemtools/iepdtd/display/container.iepd?ref=-6kRpaB0tyY>.
- [54] MCCA. Findings and recommendations of the suspicious activity report (sar) support and implementation project. Technical report, Major Cities Chiefs Association, 2008. URL <https://online.wsj.com/public/resources/documents/mccarecommendation-06132008.pdf>.

- [55] MongoDB. Production cluster architecture, May 2015. URL <https://docs.mongodb.org/manual/core/sharded-cluster-architectures-production/>.
- [56] Eileen McNulty. Sql vs. nosql- what you need to know. <http://dataconomy.com/sql-vs-nosql-need-know/>, 2014. URL <http://dataconomy.com/sql-vs-nosql-need-know/>. [Online; accessed March 2016].
- [57] C Suarez, F Ward, and J Finsterwald. Mongodb vs. sql server’s xml data type, March 2014. URL <http://lifeinvistaprint.com/techblog/mongodb-vs-sql-servers-xml-data-type/>.
- [58] Michel Mayrand. Maritime situational awareness research infrastructure (msari): Performance analysis, monitoring and optimization. Technical report, DRDC CAtlantic, March 2014.
- [59] A Isenor, M-O St-Hilaire, S Webb, and M Mayrand. Msari: A database for large volume storage and utilization of maritime data. *Journal of Navigation*, 2016.

This page is intentionally left blank.

Appendix A

Use cases

This section presents the different use cases used to determine the requirements of SARA (presented in section 5) to support the proposed design (described in section 6).

This section is organised as follows:

- Section A.1 presents a reporting anomalies use case.
- Section A.2 presents a querying anomalies use case.
- Section A.3 presents a querying ship profiles use case.
- Section A.4 presents a ship profiles edition use case.
- Section A.5 presents a user creation use case.
- Section A.6 presents a user profile edition use case.

A.1 Uploading a Ship Anomaly

Summary	A watch officer finds two ships displaying the same MMSI and identifies the one broadcasting the wrong ID. To record the anomaly, the officer uses SARA by filling the required information of an anomaly report. Upon reception, the system validates and saves the information then links the ship specified in the anomaly to an existing ship profile (when available).
----------------	---

Prerequisites	<ul style="list-style-type: none">• The user must have an account with SARA and be logged in.• The user must know the required field structure of the anomaly report. Alternatively, the 3rd party application the user employs to submit the anomaly must enforce the proper anomaly report structure.
Main scenario	<ol style="list-style-type: none">1. The user identifies a ship anomaly to store and persist.2. The user fills the required fields for the ship anomaly report and submits.3. The system validates the report contents.4. The system returns a successful transaction message.5. The system database persists the anomaly report.
Outcomes on success	The anomaly report is persisted in the database. A successful transaction message is returned to the user.

Additional outcomes or alternatives	<ol style="list-style-type: none"> 1. If the anomaly report fails to pass validation, either because of missing fields or because unexpected inputs are detected, then: <ol style="list-style-type: none"> (a) The reason of failure is identified (Eg. Incorrect Format) if possible. (b) The report is discarded. (c) A failure message is sent to the user containing the identified error and concerned field. 2. If a known ship is referenced in a successfully transmitted anomaly report: <ol style="list-style-type: none"> (a) The corresponding ship profile is marked for update at the next availability. The ship's reliability rating is to be updated according to the severity of the report and previous occurrence of the anomaly. 3. If a ship unknown to SARA is referenced in a successfully transmitted anomaly report: <ol style="list-style-type: none"> (a) A new ship profile is created, if enough information exists to satisfy the creation criteria. (b) The ship's reliability rating is scheduled to be updated according to the severity of the report. 4. If no ship is referenced in the anomaly report: <ol style="list-style-type: none"> (a) No additional action is required from the system. The report is accepted and persisted for future reference with an unknown ship. 5. If the user is not logged in: <ol style="list-style-type: none"> (a) Connection lost or other system failure (non-human error) (b) The anomaly observed is not available (not part of the taxonomy)
--	---

Table A.1: Description of the ship anomaly upload use case.

A.2 Retrieving Ship Anomalies from SARA

Summary	A watch officer is tasked to survey a region outside of Halifax harbor out to Devil's Island. After determining the coordinates delimiting the region of interest, the watch officer enters them in a SARA query along with a time period of interest. SARA returns all ship anomalies for the given time period and region.
----------------	--

Prerequisites	<ul style="list-style-type: none"> • The user must have an account with SARA and be logged in. • The user must know the required field structure of the query. Alternatively, the 3rd party application the user employs to submit the query must enforce the proper structure.
Main scenario	<ol style="list-style-type: none"> 1. The user identifies a time period and region of interest. 2. The user then enters all coordinates and time limits in a query to SARA and submits the query. 3. The system validates the query and fetches the result. 4. The system validates the report contents. 5. The system returns a successful transaction message along with the result set.
Outcomes on success	All anomaly reports within the specified time period and region of interest are returned to the user.
Additional outcomes or alternatives	<ol style="list-style-type: none"> 1. If the query fails validation, then: <ol style="list-style-type: none"> (a) The reason of failure is identified (Eg. Incorrect Format) if possible. (b) A failure message is sent to the user containing the identified error and concerned field. 2. If filters are too restrictive and no anomaly is returned. 3. Connection is lost or other system failure (non-human error). 4. Anomalies associated to unknown ships are returned. 5. Filters are not restrictive enough and the result set size exceeds the maximum threshold.

Table A.2: Description of the request ship anomaly use case.

A.3 Retrieving Ship Profiles from SARA

Summary	A SARA Ruling Authority is tasked to solve ship ambiguities occurring in SARA's database. After determining, either manually or using an automatically generated list, the ambiguous ship candidates, he sends a query to SARA. SARA returns all ship profiles requested by the Ruling Authority for further inspection and possible disambiguation.
Prerequisites	<ul style="list-style-type: none"> • The user must have a higher class account with SARA and be logged in. • The user must know the possible ambiguous ship profiles or have access to them via an automatically generated list.
Main scenario	<ol style="list-style-type: none"> 1. The user identifies the ambiguous ship profiles of interest. 2. The user then enters all the necessary details in a query to SARA and submits the query. 3. The system validates the query and fetches the result. 4. The system validates the profile contents. 5. The system returns a successful transaction message along with the result set. 6. Built-in SARA web client displays the profiles to the ruling authority for inspection and/or disambiguation.
Outcomes on success	All ship profiles of interest are returned to the user.
Additional outcomes or alternatives	<ol style="list-style-type: none"> 1. If the query fails validation. <ol style="list-style-type: none"> (a) The reason of failure is identified (Eg. Incorrect Format) if possible. (b) A failure message is sent to the user containing the identified error and concerned field. 2. Connection lost or other system failure (non-human error).

Table A.3: Description of the request ship profile use case.

A.4 Editing a Ship Profile for SARA

Summary	A SARA Ruling Authority is tasked to solve ship ambiguities occurring in SARA's database. After querying the profiles of ambiguous ships from SARA, he proceeds to their inspection. The Ruling Authority decides to merge two ship profiles by selecting the correct attributes from the ship profiles list and requesting that SARA merges them.
Prerequisites	<ul style="list-style-type: none">• The Ruling Authority must have an account with SARA and be logged in.• The Ruling Authority must have access to a web service interface for ship profile edition.
Main scenario	<ol style="list-style-type: none">1. The Ruling Authority inspects the ambiguous ship profiles.2. The Ruling Authority then corrects the profiles and sends the appropriate query to SARA to merge them in a appropriate way, so that an ambiguity is removed.3. The system validates the query and proceeds with the merging of the profile.4. The system returns a successful transaction message along with the result set.
Outcomes on success	The ship profile has been properly updated and at least one ambiguity in the system has been removed.

Additional outcomes or alternatives	<ol style="list-style-type: none"> 1. If the query fails validation. <ol style="list-style-type: none"> (a) The reason of failure is identified, if possible. (b) A failure message is sent to the user containing the identified error and concerned field.
--	--

Table A.4: Description of the ship profile edition use case.

A.5 Creating a New User Profile in SARA

Summary	A SARA system administrator is tasked with the creation of a new user profile. The administrator sends all the required info for a new user as a SARA query. The required information includes the user name, organization and access level. The system receives the query, creates the user and returns a temporary password associated to the account.
Prerequisites	<ul style="list-style-type: none"> • The administrator must be logged in. • The administrator must have access to the built-in web interface for user creation.
Main scenario	<ol style="list-style-type: none"> 1. The admin enters all information required for the creation of a new account in the form of a query to SARA. 2. The system receives and validates the information ensuring the data is correct. 3. The system creates a new account. 4. The system returns a successful transaction message. 5. The user is sent a temporary password.
Outcomes on success	A new user is created from the submitted information. The user is transmitted a temporary password and has full access to his/her account.

Additional outcomes or alternatives	<ol style="list-style-type: none"> 1. If the query fails validation: <ol style="list-style-type: none"> (a) The reason of failure is identified (Eg. Incorrect Format) if possible. (b) A failure message is sent to the user containing the identified error and concerned field. 2. If the query is sent from an account with insufficient rights for user creation: <ol style="list-style-type: none"> (a) A failure message is returned warning the sender of insufficient rights to perform the desired action. 3. The user already exists in the system (probably goes under (a)).
--	--

Table A.5: Description of the user profile creation use case.

A.6 Editing an User Profile

Summary	A SARA Administrator is tasked to modify existing user profiles in SARA's database. He sends a query to SARA to retrieve the user profiles. SARA returns all user profiles requested by the Administrator and displays it using its web interface for edition. After modification, the profile modification query is sent to SARA and properly executed.
Prerequisites	<ul style="list-style-type: none"> • The administrator must be logged in. • The administrator has access to an administration user interface for user profile modification.

Main scenario	<ol style="list-style-type: none"> 1. The administrator receives user profiles to modify. 2. The user then enters all the necessary details in a query to SARA and submits the query. 3. The system validates the query and fetches the result. 4. The system validates the profile contents. 5. The system returns a successful transaction message along with the result set. 6. Built-in SARA web client displays the profiles to enable modification of any fields. 7. After fields modification, the administrator sends the modified profile to SARA. 8. The system validates the profile contents. 9. The system returns a successful transaction message.
Outcomes on success	All user profiles of interest are modified as requested.
Additional outcomes or alternatives	<ol style="list-style-type: none"> 1. If the query fails validation. <ol style="list-style-type: none"> (a) The reason of failure is identified (Eg. Incorrect Format) if possible. (b) A failure message is sent to the user containing the identified error and concerned field. 2. Connection lost or other system failure (non-human error).

Table A.6: Description of the user profile edition use case.

This page is intentionally left blank.

Appendix B

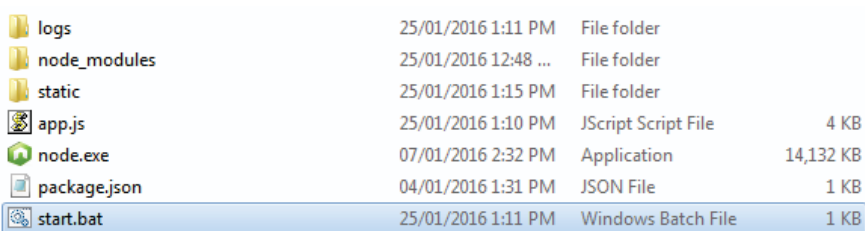
Demonstration Application Installation

This section presents the instructions to install and launch the demonstration application. It also provides solutions to problems that may happen during installation.

B.1 Installation and Launch

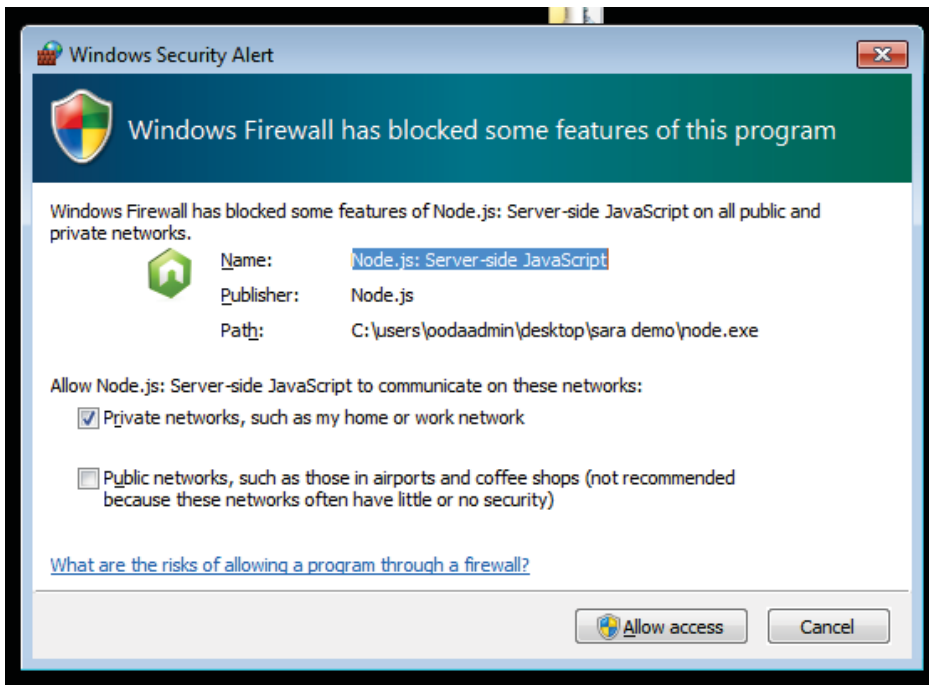
This installation assumes a Windows OS and that Chrome or Firefox is up and running.

1. Extract SARA Demo zip file.
2. Open the SARA Demo folder.

3. 

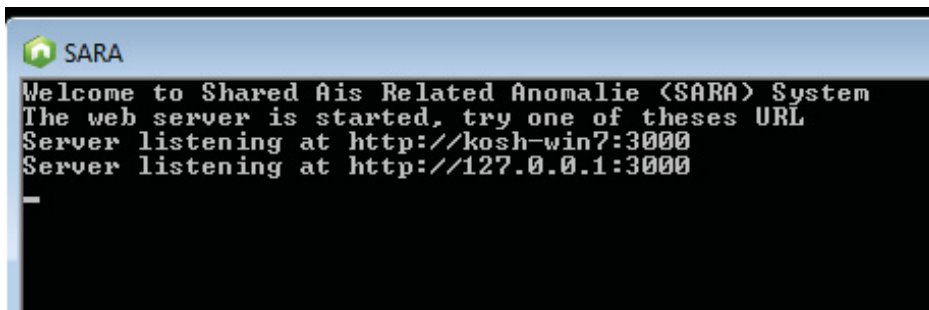
Once in the folder, double click on `start.bat`.

4.



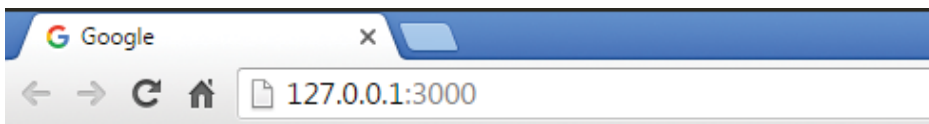
The first time the program is executed, Windows firewall will probably send an alert to notify the user that the program needs certain access. The access needs to be granted. It allows the software to open a port for the server part of the application. See section B.2 if this operation is blocked.

5.



Once started, the application will open a terminal and will display that the server is listening to at least 2 URLs.

6.



Once the server is open, open a web browser (use Chrome or Firefox) and enter one of the addresses displayed by the terminal.

7. The web application will load in few seconds and will be ready once vessels appear on screen.

B.2 Troubleshooting

1. The package contains a `.exe` and a `.bat` file that may be blocked by mail provider, firewall or browser. Renaming the extension of the file (`zip` to `zipX`) may fool the scanners mentioned. Then renaming the extension back to `zip` (`zipX` to `zip`) should work.
2. When executing for the first time, you may need to be an administrator to accept the new firewall rule. As a consequence, a system administrator may be required to execute the application for the first time.
3. The server application needs `nodejs` (free software) to run. The executable provided is should run on a x64 operating system. If you are running a x86 operating system, you may need to download a 32 bits executable of `nodejs` version 4.2.6. (available at <https://nodejs.org/dist/v4.2.6/win-x86/node.exe>).